

A hierarchical state space approach to affective dynamics

Tom Lodewyckx^{a,*}, Francis Tuerlinckx^a, Peter Kuppens^{a,b}, Nicholas B. Allen^b, Lisa Sheeber^c

^a University of Leuven, Belgium

^b University of Melbourne, Australia

^c Oregon Research Institute, United States

ARTICLE INFO

Article history:

Received 4 February 2010

Received in revised form

27 July 2010

Available online 20 October 2010

Keywords:

Bayesian hierarchical modeling

Cardiovascular processes

Emotions

Linear dynamical system

State space modeling

ABSTRACT

Linear dynamical system theory is a broad theoretical framework that has been applied in various research areas such as engineering, econometrics and recently in psychology. It quantifies the relations between observed inputs and outputs that are connected through a set of latent state variables. State space models are used to investigate the dynamical properties of these latent quantities. These models are especially of interest in the study of emotion dynamics, with the system representing the evolving emotion components of an individual. However, for simultaneous modeling of individual and population differences, a hierarchical extension of the basic state space model is necessary. Therefore, we introduce a Bayesian hierarchical model with random effects for the system parameters. Further, we apply our model to data that were collected using the Oregon adolescent interaction task: 66 normal and 67 depressed adolescents engaged in a conflict-oriented interaction with their parents and second-to-second physiological and behavioral measures were obtained. System parameters in normal and depressed adolescents were compared, which led to interesting discussions in the light of findings in recent literature on the links between cardiovascular processes, emotion dynamics and depression. We illustrate that our approach is flexible and general: The model can be applied to any time series for multiple systems (where a system can represent any entity) and moreover, one is free to focus on various components of this versatile model.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

The role affect and emotions play in our daily life can hardly be overestimated. Affect and emotions produce the highs and lows of our lives. We are happy when a paper gets accepted, we are angry if a colleague intentionally spreads gossip about us and we feel guilty when we cross a deadline for a review. For some people, their affect is a continuous source of trouble because they suffer from affective disorders, such as a specific phobia or depression.

As for any aspect of human behavior, emotions are extremely complex phenomena, for several reasons. First, they are multicomponential, consisting of experiential, physiological and behavioral components (Gross, 2002). If you are afraid when walking alone on a deserted street late at night, this may lead to bodily effects such as heart palpitations but also to misperceptions of stimuli in the environment and a tendency to walk faster. Second, an emotion fluctuates over time. Without going into the muddy waters of what the exact definition of an emotion is (see Bradley & Lang, 2007), it is clear that emotions function to signal relevant events

for our goals (Oatley & Jenkins, 1992). Because of their communicative function, emotions have a clear temporal component (see e.g., Frijda, Mesquita, Sonnemans, & Van Goozen, 1991) and therefore a genuine understanding of emotions implies an understanding of their underlying dynamics. Third, emotional reactions are subject to contextual and individual differences (see e.g., Barrett, Mesquita, Ochsner, & Gross, 2007; Kuppens, Stouten, & Mesquita, 2009).

These complicating factors make the study of affective dynamics a challenging research domain, which requires an understanding of the complex interplay between the different emotion components across time, context and individual differences (Scherer, 2000, 2009). Despite its importance and complexity, research on emotion dynamics is still in its infancy (Scherer, 2000). Aside from the lack of a definite theoretical understanding of affective phenomena, a large part of the reason for this lies in the complexity of the data involved in such an enterprise. For instance, because of the prominent physiological component of affect, biological signal processing techniques are required and these are typically not part of the psychology curriculum. On the other hand, the existing methods, traditionally developed and studied in the engineering science, are not directly applicable. As we explain below, we believe one of the major bottlenecks is the existence of individual differences. Indeed, as noted by Davidson (1998), one of the most striking features of emotions is the presence of significant individual differences in almost all aspects involved in their

* Corresponding address: University of Leuven, Department of Psychology, Tiensestraat 102 Box 3713, B-3000 Leuven, Belgium.

E-mail address: tom.lodewyckx@psy.kuleuven.be (T. Lodewyckx).

elicitation and control. Incorporating such differences is therefore crucial, not only to account fully for the entire range of emotion dynamics across individuals, but also for studying the differences between individuals characterized by adaptive or maladaptive emotional functioning.

As an example, let us introduce the data that will be discussed and investigated below. Two groups of adolescents (one with Major Depressive Disorder and the other without any emotional or behavioral problems) engaged in an interaction task with their parents during a few minutes in which they discussed and tried to resolve a topic of conflict. During the task, several physiological measures were recorded from the adolescent. Moreover, the behavior of the adolescent and parents was observed and micro-socially coded. All measures were obtained on a second-to-second basis. Several possible research questions are: In what way do the physiological dynamics differ between depressed adolescents and healthy controls (hereafter referred to as normals)? What is the effect of the display of angry behavior by a parent on the affective physiology observed in the adolescent, and is this effect different for depressed and normal adolescents?

A powerful modeling framework that is capable for addressing the above questions is provided by state space modeling. State space models will be explained in detail in the next section, but for now it suffices to say that they have been developed to model the dynamics of a system from measured inputs and outputs using latent states. Usually, it is a single system that is being studied. However, in the particular example in this paper there are as many systems as participants. Because a single state space is already a complex model for statistical inference, studying several of these state space models simultaneously is a daunting task. However, in the present paper we offer a solution to this problem by incorporating state space models in a hierarchical Bayesian framework, which allows one to study multiple systems (e.g., individuals) simultaneously, and thus allows one to make inferences about differences between individuals in terms of their affective dynamics. Markov chain Monte Carlo methods make the task of statistical inference for such hierarchical models more digestible and the Bayesian approach lets us summarize the most important findings in a straightforward way.

In sum, the goal of this paper is to introduce a hierarchical state space framework allowing us to study individual differences in the dynamics of the affective system. The outline of the paper is as follows. In the next section, we introduce a particular state space model, the linear Gaussian state space model, and extend it to a hierarchical model. In subsequent section, we illustrate the framework by applying it to data consisting of cardiovascular and behavioral measures that were taken during the interaction study introduced above. By focusing on various aspects of the model, we will show how our approach allows us to explore several research questions and address specific hypotheses that are discussed in the literature on the physiology and dynamics of emotions. Finally, the discussion reviews the strengths and limitations of the model, and we make some suggestions for future developments.

2. Hierarchical state space modeling

We will present a model for the affective dynamics of a single person and then extend this model hierarchically: The basic model structure for each person's dynamics will be of the same type, and the hierarchical nature of the model allows for the key parameters of the model to differ across individuals.

The single individual's model will be a linear dynamical systems model, cast in the state space framework. In the next paragraphs, the key concepts will be explained verbally and in the next subsections, the mathematical aspects of the model will be described. Our coverage of dynamical systems theory is intended

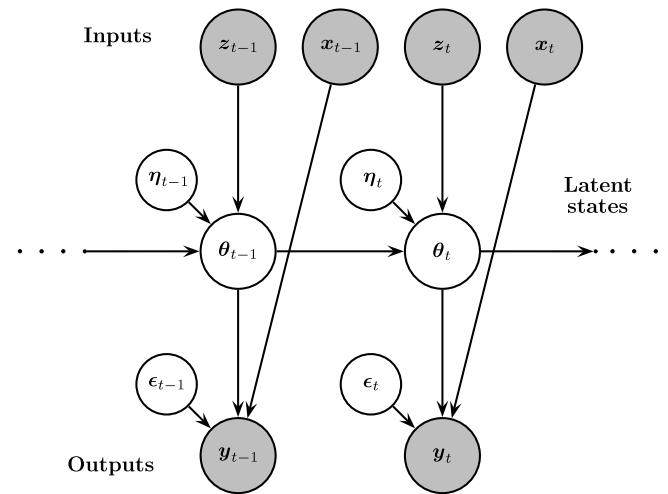


Fig. 1. Graphical representation of a linear dynamical system.

to be sufficient to understand the remainder of the paper, but readers who wish a deeper study of the topic can consult [Grewal and Andrews \(2001\)](#), [Ljung \(1999\)](#) or [Simon \(2006\)](#). Suppose we have a set of observable output variables measured at time t ($t = 1, 2, \dots, T$) collected in a vector y_t (e.g., heart rate and blood pressure measured every second from an adolescent) and a set of observable input variables (e.g., an indicator coding whether a parent is angry or not), collected in the vectors x_t and z_t , that are believed to have, respectively, a direct and indirect influence on the outputs. It is believed that the input and output are part of a system, which is a mathematical description of how inputs affect the outputs through state variables (denoted by θ_t). The state variables are a set of latent variables of the system that exhibit the dynamics. A graphical illustration of a state space system is given in [Fig. 1](#). In this paper, we will work in discrete time (as opposed to continuous time, for which $t \in \mathbb{R}$). The dynamics of the states are then described by a stochastic difference equation where the stochastic term is called the process noise or the innovation, denoted by η_t . As a result, θ_t cannot be perfectly predicted from θ_{t-1} and z_t . On the observed level, it is assumed that the measurements of the output variables are noisy and the measurement error at time t is denoted by ϵ_t .

The state space modeling framework has been applied in a wide variety of scientific disciplines: From the Apollo space program to forecasting in economical time series to robotics and ecology. The important breakthrough for the practical application of state space models was the seminal paper by [Kalman \(1960\)](#) on the estimation of the states in a linear model.¹ The paper introduces what became known as the Kalman filter, which is a recursive least squares estimator of the state vector θ_t at time t given all observations up to time t (i.e., y_1, \dots, y_t). Note that the Kalman filter can be derived from a Bayesian argumentation (see [Meinhold & Singpurwalla, 1983](#), see also below).

The Kalman filter is used in state space modeling to obtain an estimate of the latent states, but depending on the context, the researcher may or may not be interested in the values of these states. As an example of the former attitude let us consider an aerospace application. Here the latent states can be the position coordinates of a spacecraft and the parameters of the dynamical system are entirely known (because they may be governed by well-known Newtonian mechanics). In such case, the question of

¹ Interestingly, the same idea was described already by the Danish statistician Thorvald Thiele in 1880, in which he also introduces Brownian motion (see [Lauritzen, 1981](#)).

interest is to determine where the spacecraft is, and if necessary, to correct its trajectory (i.e., exerting control by changing the input). However, in other application settings, the dynamics of the process are not exactly known and the interest lies in inference on the parameters of the model. Estimating the states, for instance through the Kalman filter, is only a necessary step in the process of inferring the dynamics of the system. This is the topic of system identification (Ljung, 1999): constructing a mathematical dynamical model based on the noisy measurements.

The specific type of state space model that will be used to model the dynamics of a single individual is a linear state space model: The state dynamics are described using a linear difference equation (with the random process noise term added). Although the kind of processes we will model may be considered to be inherently nonlinear, linear dynamical systems often serve as a good approximation that may reveal many aspects of the data, even when it is known that the underlying system is nonlinear (Van Overschee & De Moor, 1996).

In the following subsections, we will explain in more detail the linear Gaussian state space model and then discuss the hierarchical extension.

2.1. The linear Gaussian state space model

The state space model consists of two equations: A transition (or state) equation and an observation equation. Let us start with the transition equation that represents the dynamics of the system. As defined above, θ_t is the vector (of length P) of latent state values at time t . In addition, suppose there are K state covariate measurements collected in a vector \mathbf{z}_t . The type of transition equation we use in this paper can be written as follows:

$$\theta_t | \theta_{t-1}, \dots, \theta_{t-L} \sim N \left(\sum_{l=1}^L \Phi_l \theta_{t-l} + \Delta \mathbf{z}_t, \Sigma_\eta \right). \quad (1)$$

From Eq. (1), it can be seen that the P -dimensional vector of latent processes θ_t is modeled as the combination of a vector autoregression model of order L (denoted as VAR(L)) and a multivariate regression model.² The order of the vector autoregressive part is the maximum lag L . The transition matrix Φ_l of dimension $P \times P$ (for $l = 1, \dots, L$) consists of autoregression coefficients on the diagonal and crossregression coefficients in the off-diagonal slots. The autoregression coefficient $\phi_{[a,a]l}$ represents the dependence of latent state $\theta_{[a]t}$ on itself l time points before, that is, the effect of state $\theta_{[a]t-l}$. The crossregression coefficient $\phi_{[a,b]l}$ represents the influence of state $\theta_{[b]t-l}$ on $\theta_{[a]t}$. The regression coefficient matrix Δ of dimension $P \times K$ contains the effects of the K covariates in \mathbf{z}_t on the latent processes θ_t . The innovation covariance matrix Σ_η of dimension $P \times P$ describes the Gaussian fluctuations and covariations of the P -dimensional innovation vectors η_t .

The model presented in Fig. 1 corresponds to a model for which $L = 1$ (i.e., VAR(1)) in Eq. (1). However, if we ignore the contribution of the state covariates in Eq. (1), the general VAR(L) process can also be defined as a VAR(1) process as follows:

$$\theta_t^* | \theta_{t-1}^* \sim N (\mathbf{F} \theta_{t-1}^*, \mathbf{e}_1 \mathbf{e}_1' \otimes \Sigma_\eta), \quad (2)$$

where the definition of the new state vector is: $\theta_t^* = (\theta_t', \dots, \theta_{t-L}')'$, or in words, the non-lagged and lagged state vectors combined in a new state vector θ_t^* . The symbol \otimes represents the

Kronecker product.³ The new transition matrix \mathbf{F} of dimension $PL \times PL$ is equal to

$$\mathbf{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{L-1} & \Phi_L \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix}, \quad (3)$$

where \mathbf{I} represents a $P \times P$ identity matrix and $\mathbf{0}$ refers to a $P \times P$ matrix of zeros. This technique of reformulating a dynamical system of order L into one of order 1 is common in state space modeling, but is also used in differential equation modeling (see Borrelli & Coleman, 1998). Because in our new formulation, some parts of θ_t^* and θ_{t-1}^* are exactly equal to each other (there is no randomness involved), the covariance matrix in Eq. (2) is degenerate:

$$\mathbf{e}_1 \mathbf{e}_1' \otimes \Sigma_\eta = \begin{bmatrix} \Sigma_\eta & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}, \quad (4)$$

where \mathbf{e}_1 is the L -dimensional first standard unit vector (with 1 in the first position and zero otherwise). Eq. (2) shows that it is most important to study the VAR(1) process since higher order processes can always be reduced to a lag 1-model. For instance, a useful result is that if the process in Eq. (2) is stationary, all eigenvalues of \mathbf{F} should lie inside the unit circle (see e.g., Hamilton, 1994). The state covariate regression part can be easily added to Eq. (2) by inserting in the mean structure: $\mathbf{e}_1 \otimes \Delta \mathbf{z}_t$.

The second equation in a state space model is the observation equation, which reads for the model as presented in Eq. (1) as follows:

$$\mathbf{y}_t | \theta_t \sim N (\boldsymbol{\mu} + \Psi \theta_t + \Gamma \mathbf{x}_t, \Sigma_\epsilon), \quad (5)$$

where the observed output at time t is collected in a Q -dimensional vector \mathbf{y}_t . The observation equation maps the vector of latent states θ_t at time t to the observed output vector \mathbf{y}_t also at time t (hence there are no dynamics in this equation). The observation equation also contains a P -dimensional mean vector $\boldsymbol{\mu}$ and a $Q \times P$ design matrix Ψ that specifies to which extent each of the latent states influence the observed output processes. In addition, there is the influence of the J observation covariates collected in a vector \mathbf{x}_t , with the corresponding regression coefficients being organized in the $Q \times J$ matrix Γ . The $Q \times Q$ measurement error covariance matrix Σ_ϵ describes the Gaussian fluctuations and covariations of the measurement errors (denoted as ϵ_t).

It should be noted that in this paper we consider a restricted version of the state space model in which each observed output is tied directly to a latent state. This means that $P = Q$ and $\Psi = \mathbf{I}_P$. Stated otherwise, each observed process has a corresponding latent process.

If the model from Eq. (2) is used as the transition equation, then the observation equation needs to be modified as well because the definition of the latent states has changed:

$$\mathbf{y}_t | \theta_t^* \sim N (\boldsymbol{\mu} + (\mathbf{e}_1' \otimes \Psi) \theta_t^* + \Gamma \mathbf{x}_t, \Sigma_\epsilon), \quad (6)$$

where \mathbf{e}_1 is again the first standard unit vector.

As mentioned before, the graphical representation in Fig. 1 is in fact much more general than it appears. It seems to be only valid for a restricted dynamical model of only lag 1, but as we have shown

² Although we focus on (vector) autoregressive processes in the state equation, the general state space framework is more general and can allow for vector autoregressive moving average (VARMA) processes in the state equation (see e.g., Durbin & Koopman, 2001).

³ The Kronecker product $A \otimes B$ of $m \times n$ matrix A and $p \times q$ matrix B is a $mp \times nq$ matrix consisting of $m \times n$ block matrices each of size $p \times q$ such that the block (i, j) is defined as $a_{ij}B$, where a_{ij} is element in position (i, j) of matrix A .

the lag 1 dynamical model also has the lag L model as a special case after an appropriate definition of the state vector. Fig. 1 is not only an appealing illustration of some of the general components of the model; it also gives a concise and detailed description of the basic properties of the model. Graphical modeling is a popular method for visualizing conditional dependency relations between the various components of the model. For instance, one can read off quite easily that the Markov property holds for the transition equation on the latent state level: Using the knowledge of state θ_t , there is independence between θ_{t-1} and θ_{t+1} . In the language of graphical models (see e.g., Pearl, 2000), this means that θ_{t-1} and θ_{t+1} are d -separated by θ_t (i.e., θ_t blocks the path from θ_{t-1} to θ_{t+1}) and therefore θ_{t-1} and θ_{t+1} are conditionally independent given θ_t . Using the same reasoning, it can be seen that there is no conditional independence between y_{t-1} and y_{t+1} given y_t because the latter one does not block the path from y_{t-1} to y_{t+1} . Hence, while the dynamical properties are limited to the latent dynamical system and they are Markovian, the dependency relations between the observed outputs y_t are not Markovian but more complex.

2.2. Hierarchical extension

In the previous subsection we have described a linear dynamical model for a single system. In many scientific domains one can stop there and discuss how the statistical inference for such a single-system linear dynamical model can be performed. However, in psychology, if the system refers to processes within one individual, the study of individual differences in such processes requires studying many systems and the differences between them simultaneously. There are several options for doing this.

A first option is to analyze the data from all subjects simultaneously. However, such an approach is hard to reconcile with the existence of individual differences unless a model structure is chosen that allows for states to be unique for different individuals. A disadvantage of such an approach is that the state vector quickly becomes very large and the computations unwieldy. A second option is to apply a separate linear dynamical model to the data of each participant. Although this approach allows for maximal flexibility in terms of how people may differ from each other, it is not without risk. In many cases, data from single participants are not very informative. In the example to be discussed in detail later on, the anger responses of the parents will be considered as input for the cardiovascular system. Unfortunately, many parents show angry behavior only at rare occasions, which yields uninformative data and therefore may hamper the inferential process.

An interesting modeling approach that allows for individual differences but at the same time does not collapse under the weight of its flexibility is hierarchical linear dynamical modeling. In hierarchical models, person-specific key parameters are assumed to be a sample from a distribution, typical for the population that is investigated, allowing us to model both individual differences and group effects (Lee & Webb, 2005; Shiffrin, Lee, Kim, & Wagenmakers, 2008). The same approach to modeling individual differences is used throughout this special issue, including in modeling memory (Pratte & Rouder, 2011), confidence (Merkle, Smithson, & Verkuilen, 2011) and decision-making (Nilsson, Rieskamp, & Wagenmakers, 2011).

To explain the hierarchical extension, let us first reconsider the state space model and make it specific for person i (with $i = 1, \dots, I$):

$$y_{t,i} | \theta_{t,i} \sim N(\mu_i + \theta_{t,i} + \Gamma_i x_{t,i}, \Sigma_\epsilon)$$

$$\theta_{t,i} | \theta_{t-1,i}, \dots, \theta_{t-L,i} \sim N\left(\sum_{l=1}^L \Phi_{l,i} \theta_{t-l,i} + \Delta_i z_{t,i}, \Sigma_\eta\right), \quad (7)$$

such that all parameters are specific for individual i , except the error covariance matrices Σ_η and Σ_ϵ . Although it would be realistic, we have chosen not to allow these covariance matrices to vary between individuals because hierarchical models for covariance matrices are not yet very well understood (for an exception, see Pourahmadi & Daniels, 2002).

The size of the vectors and matrices in Eq. (7) is the same as in the corresponding Eqs. (1) and (5). In order to have a parsimonious notation scheme, we will denote elements of the person-specific vectors and matrices with an index i and the appropriate row and column indicators between square brackets. For instance, the p th element of μ_i is then equal to $\mu_{[p]i}$ and the element (j, k) of $\Phi_{l,i}$ is $\phi_{[j,k]l,i}$.

A next step in constructing a hierarchical model is to specify the population distributions from which it is assumed that the person-specific parameters are sampled. For each element of these system parameters, an independent hierarchical multivariate normal distribution is assumed. For notational convenience, the population mean is always denoted by α (mean) or α (mean vector) and the population variance by β (variance) or B (covariance matrix). An appropriate subscript will be used to specify for which particular parameter hierarchical parameters are defined. For example, the person-specific mean μ_i is assumed to be sampled from a multivariate normal distribution with mean vector α_μ and covariance matrix B_μ :

$$\mu_i \sim N(\alpha_\mu, B_\mu). \quad (8)$$

The other person-specific parameters are all organized into matrices and to specify their corresponding population distributions, we need the vectorization operator $\text{vec}(\cdot)$ (or vec operator; see Magnus & Neudecker, 1999), which transforms a matrix into a column vector by reading the elements row-wise and stacking them below one another. Let us illustrate this for the transition matrix $\Phi_{1,i}$. Assuming that $P = 2$ (i.e., there are two latent states), $\Phi_{1,i}$ is a 2×2 matrix that is vectorized as follows:

$$\text{vec}(\Phi_{1,i}) = \text{vec}\left(\begin{bmatrix} \phi_{[1,1]1,i} & \phi_{[1,2]1,i} \\ \phi_{[2,1]1,i} & \phi_{[2,2]1,i} \end{bmatrix}\right) = \begin{bmatrix} \phi_{[1,1]1,i} \\ \phi_{[1,2]1,i} \\ \phi_{[2,1]1,i} \\ \phi_{[2,2]1,i} \end{bmatrix}. \quad (9)$$

As for the hierarchical distributions for these matrix parameters $\Phi_{l,i}$ (for $l = 1, \dots, L$), Γ_i and Δ_i , we assume multivariate normal distributions for all vectorized parameters. For the regression coefficients of the observation covariates $x_{t,i}$, the population distribution reads as

$$\text{vec}(\Gamma'_i) \sim N(\alpha_\Gamma, B_\Gamma) \quad (10)$$

and for the regression coefficients of the state covariates $z_{t,i}$, it is

$$\text{vec}(\Delta'_i) \sim N(\alpha_\Delta, B_\Delta). \quad (11)$$

For the transition matrices $\Phi_{l,i}$ (with $l = 1, \dots, L$), it is in principle the same but with the additional constraint that the dynamic latent process is stationary (i.e., that the eigenvalues of the F_i matrix fall inside the unit circle). This constraint is denoted as $I(|\lambda(F_i)| < 1)$ in subscript of the multivariate normal distribution:

$$\begin{pmatrix} \text{vec}(\Phi'_{1,i}) \\ \vdots \\ \text{vec}(\Phi'_{L,i}) \end{pmatrix} \sim N(\alpha_\Phi, B_\Phi)_{I(|\lambda(F_i)| < 1)}. \quad (12)$$

The covariance matrices can be unstructured (i.e., all variances and covariances can be estimated freely) but in the application section we will impose some structure (because it is the first time the model has been formulated and we do not want to run the risk of having weakly identified parameters that cause trouble for the convergence and mixing of the MCMC algorithm). In this paper, it is assumed that all person-specific parameters have a separate variance, but that all covariances are zero.

2.3. Statistical inference

In this section we will outline how to perform the statistical inference for the hierarchical linear dynamical model, i.e. how to estimate the model parameters with Bayesian statistics. Concerning the estimation of the parameters of a (non-hierarchical) linear dynamical model, there is a huge literature mainly in engineering (e.g., Ljung, 1999) and econometrics (e.g., Hamilton, 1994; Kim & Nelson, 1999), but also in statistics (e.g., Petris, Petrone, & Campagnoli, 2009; Shumway & Stoffer, 2006). However, to the best of our knowledge, the hierarchical extension as presented in this paper is new and estimating the model's parameters requires some specific choices, of which opting for the Bayesian approach is the most consequential one.⁴

Besides the theoretical appeal of Bayesian statistics, it also offers the possibility to use sampling based computation methods such as Markov chain Monte Carlo (MCMC, see e.g., Gilks, Richardson, & Spiegelhalter, 1996; Robert & Casella, 2004). Especially for hierarchical models this is very helpful because the multidimensional integration against the random effects densities may not have closed form solutions and if the latter do not exist, numerical procedures are practically not feasible. For instance, in the model defined in the preceding section, an integral of very high dimension has to be calculated: $P + LP^2 + PK + QJ$, where the contributions come from respectively the mean vector, the transition matrices, the state covariate regression weights, and the observation covariate regression weights. Even for a model with no covariates, two states and five time lags, the dimension is already 22. For this reason, we resort to Bayesian statistical inference and MCMC.

To sample the parameters from the posterior distribution, we have implemented a Gibbs sampler (Casella & George, 1992; Gelfand & Smith, 1990). In this method, the vector of all parameters is subdivided into blocks (in the most extreme version, each block consists of a single parameter) and then we sample alternatively from each full conditional distribution: The conditional distribution of a block, given all other blocks and the data. For our model, all full conditionals are known and easy-to-sample distributions (i.e., normals, scaled inverse chi-squares). Sampling iteratively from the full conditionals creates a Markov chain with as an equilibrium or stationary distribution the posterior distribution of interest. Although strictly speaking convergence is only attained in the limit, in practice one lets the Gibbs sampler run for a sufficiently long “burn-in period” (long enough to be confident about the convergence) and after the burn-in period the samples are considered to be simulated from the true posterior distribution. Convergence can be checked with specifically developed statistics (for instance \hat{R} , see Gelman, Carlin, Stern, & Rubin, 2004). The inferences are then based on the simulated parameter values (for instance, the sample average of a set of draws for a certain parameter is taken to be a Monte Carlo estimate of the posterior mean for that parameter). More information on the theory and practice of the Gibbs sampler and other MCMC methods can be found in Gelman et al. (2004) and Robert and Casella (2004).

For describing the specific aspects of the Gibbs sampler used in this paper, we start by reconsidering the observation and transition equations, see Eq. (7). It can be seen that if the latent states $\theta_{t,i}$ were known for all time points $t = 1, \dots, T$ for a specific subject i , the problem of estimating the other parameters reduces to estimating the coefficients and residual variances in a regression model (see Gelman et al., 2004). Unfortunately, the latent states $\theta_{t,i}$ are not known, but they can be considered as

“latent data” or “missing data” and their full conditional (given the model parameters and the data) can be derived as well. Note that considering the latent states as missing data and sampling them will enormously simplify the sampling algorithm. After running the sampler, the sampled latent states may be used as well for inference or they can be ignored (and this option actually means that one integrates the latent states out of the posterior).

The Gibbs sampler consists of three distinct large components: (1) sampling the latent states, (2) sampling the system parameters, and (3) sampling the hierarchical parameters. These components will be discussed separately in the next paragraphs.

Component 1: Sampling the latent states In the first component, we estimate the latent states with the forward filtering backward sampling algorithm (FFBS; Carter & Kohn, 1994). Let us define $\Theta_{1:t,i}$ as the collection of latent state vectors of individual i up to time t . Likewise, $\mathbf{y}_{1:t,i}$ contains all observations for individual i up to time t . In addition, collect all parameters of the model in a vector ξ . The FFBS algorithm draws a sample from the conditional distribution of all states for an individual i , given all data for person i and the parameters: $p(\Theta_{1:t,i} | \mathbf{y}_{1:t,i}, \xi)$. The FFBS algorithm is also called a simulation smoother (Durbin & Koopman, 2001). To simplify notation, we will suppress the dependence upon the parameters ξ in the following paragraphs. We also note that in our derivation we will assume a lag 1 dynamical model for the states. However, if one considers a more general lag L dynamical model, then $\Theta_{1:t,i}$ has to be replaced by $\Theta_{1:t,i}^*$ (containing all latent states $\theta_{t,i}^*$ defined as in Eq. (2) up to time t).

In explaining the FFBS algorithm, it is important to note that our graphical model for a single person is a Directed Acyclic Graph (DAG; directed because all links have an arrow and the graph is acyclic because there is no path connecting a node with itself) with all specified distributions being normal (both marginal and conditional distributions). As shown in Bishop (2006), in a DAG with only normal distributions, the joint distribution $p(\Theta_{1:t,i}, \mathbf{y}_{1:t,i})$ is also (multivariate) normal. Therefore, the conditional distribution $p(\Theta_{1:t,i} | \mathbf{y}_{1:t,i})$ will also be multivariate normal.

Instead of sampling directly from $p(\Theta_{1:t,i} | \mathbf{y}_{1:t,i})$, one could use in principle samples from the full conditional of a single $\theta_{t,i}$. In that case, the full conditional of $\theta_{t,i}$ given the latent states, parameters and data is used: $p(\theta_{t,i} | \theta_{1:t-1,i}, \theta_{t+1:T,i}, \mathbf{y}_{1:T,i}) = p(\theta_{t,i} | \theta_{t-1,i}, \theta_{t+1,i}, \mathbf{y}_{t,i})$. This identity holds, because given $\theta_{t-1,i}$, $\theta_{t+1,i}$ and $\mathbf{y}_{t,i}$, $\theta_{t,i}$ is independent from all other states and data (again this can be checked in the graphical model because these three nodes block all paths to the state at time t). Carter and Kohn (1994) call this the single move sampler and note that it may lead to high autocorrelations in the sampled latent states. Therefore, we opt for the multimove sampler such that $\Theta_{1:t,i}$ is sampled as a whole from $p(\Theta_{1:t,i} | \mathbf{y}_{1:t,i})$.

However, sampling efficiently from $p(\Theta_{1:t,i} | \mathbf{y}_{1:t,i})$ is not a straightforward task because it may be a very high-dimensional multivariate normal distribution (e.g., if there are 60 time points and four latent states, it is 240-dimensional). Carter and Kohn (1994) described an efficient algorithm (see also Kim & Nelson, 1999). The basic idea is based on the following factorization (repeated from Kim & Nelson, 1999):

$$\begin{aligned} p(\Theta_{1:T,i} | \mathbf{y}_{1:T,i}) &= p(\theta_{T,i} | \mathbf{y}_{1:T,i}) p(\Theta_{1:T-1,i} | \theta_{T,i}, \mathbf{y}_{1:T,i}) \\ &= p(\theta_{T,i} | \mathbf{y}_{1:T,i}) p(\theta_{T-1,i} | \theta_{T,i}, \mathbf{y}_{1:T,i}) p(\Theta_{1:T-2,i} | \theta_{T,i}, \theta_{T-1,i}, \mathbf{y}_{1:T,i}) \\ &= p(\theta_{T,i} | \mathbf{y}_{1:T,i}) p(\theta_{T-1,i} | \theta_{T,i}, \mathbf{y}_{1:T,i}) p(\theta_{T-2,i} | \theta_{T,i}, \theta_{T-1,i}, \mathbf{y}_{1:T,i}) \\ &\quad \times p(\Theta_{1:T-3,i} | \theta_{T,i}, \theta_{T-1,i}, \theta_{T-2,i}, \mathbf{y}_{1:T,i}) \\ &= \dots \\ &= p(\theta_{T,i} | \mathbf{y}_{1:T,i}) p(\theta_{T-1,i} | \theta_{T,i}, \mathbf{y}_{1:T,i}) p(\theta_{T-2,i} | \theta_{T-1,i}, \mathbf{y}_{1:T,i}) \\ &\quad \times p(\theta_{T-3,i} | \theta_{T-2,i}, \mathbf{y}_{1:T,i}) \dots p(\theta_{1,i} | \theta_{2,i}, \mathbf{y}_{1:T,i}) \\ &= p(\theta_{T,i} | \mathbf{y}_{1:T,i}) p(\theta_{T-1,i} | \theta_{T,i}, \mathbf{y}_{1:T-1,i}) p(\theta_{T-2,i} | \theta_{T-1,i}, \mathbf{y}_{1:T-2,i}) \end{aligned}$$

⁴ We did encounter a research report by Liu, Niu, and Wu (2004), but their approach differs in several aspects from our approach (for instance, it is partly Bayesian and uses the EM algorithm for some of the routines).

$$\begin{aligned} & \times p(\theta_{T-3,i}|\theta_{T-2,i}, \mathbf{y}_{1:T-3,i}) \cdots p(\theta_{1,i}|\theta_{2,i}, \mathbf{y}_{1:1,i}) \\ & = p(\theta_{T,i}|\mathbf{y}_{1:T,i}) \prod_{t=1}^{T-1} p(\theta_{t,i}|\theta_{t+1,i}, \mathbf{y}_{1:t,i}). \end{aligned} \quad (13)$$

The factorization shows that we can start sampling $\theta_{T,i}$ given all data and then going backwards, each time conditioning on the previously sampled value and the data up to that point. It is here that the Kalman filter comes into play because $p(\theta_{T,i}|\mathbf{y}_{1:T,i})$ is the so-called filtered state density at time T : It contains all information in the data about $\theta_{T,i}$. Although we will not present the mean and covariance matrix for the normal density $p(\theta_{T,i}|\mathbf{y}_{1:T,i})$ (see e.g., Bishop, 2006; Kim & Nelson, 1999; Meinhold & Singpurwalla, 1983), it is useful to stress that the Kalman filter equations can be seen as a form of Bayesian updating with $p(\theta_{T,i}|\mathbf{y}_{1:T-1,i})$ as the prior, $p(\mathbf{y}_{T,i}|\theta_{T,i})$ as the likelihood and $p(\theta_{T,i}|\mathbf{y}_{1:T,i})$ as the posterior (all distributions are normal).

The other densities $p(\theta_{t,i}|\theta_{t+1,i}, \mathbf{y}_{1:t,i})$ involved in the factorization in Eq. (13) can also be found through applying the Kalman filter. For instance, $p(\theta_{t,i}|\theta_{t+1,i}, \mathbf{y}_{1:t,i})$ is simply the Kalman filtered density for the state at time t after having seen all data up to time t and an additional “artificial” observation $\theta_{t+1,i}$, that is, the sampled value for the next state. Hence, this involves another one-step ahead prediction using the Kalman filter.

In practice, the FFBS algorithm works by running the Kalman filter forward and thus calculating the filtered densities for the consecutive states given $\mathbf{y}_{1:t,i}$ (with $t = 1, \dots, T$). After this forward step, the backward sampling is applied and this involves each time (except for the first sampled value $\theta_{T,i}$) another one-step ahead Kalman filter prediction, starting from the observed sequence $\mathbf{y}_{1:t,i}$ and adding a new “observation” $\theta_{t+1,i}$ (with an appropriate likelihood $p(\theta_{t+1,i}|\theta_{t,i})$ which is in this case the transition equation and the prior is the filtered density $p(\theta_{t,i}|\mathbf{y}_{1:t,i})$).

Component 2: Sampling the system parameters In the second component, we sample the individual effect parameters μ_i , $\Phi_{l,i}$ (for $l = 1, \dots, L$), Γ_i , Δ_i and also the covariance matrices Σ_η and Σ_ϵ from their full conditionals. Because the sampling distributions condition on the values for the latent states, obtained in the FFBS, the observation and transition equations for individual i as in Eq. (7) reduce to multivariate linear regression equations in which the regression coefficients and the covariance matrices are to be estimated. The hierarchical distributions serve as prior distributions for the random effects. Because this is a well-known problem in Bayesian statistics, we refer the reader to the literature for more information (see e.g., Gelman et al., 2004). More information about the derivation of the full conditionals can be found in Appendix A.

Component 3: Sampling the hierarchical parameters In this component, the hierarchical mean vectors α_μ , α_ϕ , α_Γ and α_Δ and the hierarchical covariance matrices \mathbf{B}_μ , \mathbf{B}_ϕ , \mathbf{B}_Γ and \mathbf{B}_Δ are sampled. These parameters only depend on the individual effects that were sampled in the previous component. Hence, given the sampled individual system parameters, the problem reduces to estimating the mean vector and covariance matrix for the normal population distribution. Technical details are explained in Appendix A.

Easy-to-use software WinBUGS is available for free and used in various research areas as a standard in Bayesian modeling (Lunn, Thomas, Best, & Spiegelhalter, 2000). However, for optimal tuning and the implementation of complex subroutines in the MCMC sampling scheme, custom written scripts are a better alternative, certainly for complex models as the one presented here. For the hierarchical state space model, the implementation of a Gibbs sampler was especially written in R (R Development Core Team, 2009).⁵

Before turning to the application, we want to make two comments about the estimation procedure. A first observation is that when estimating the model using artificial data (simulated with known true values), we obtained estimates that were reasonably close to the population values. An illustration of this can be found in Appendix B. A second point is that the R code can be speeded up by coding some parts or the whole code in a compiled programming language (e.g., C++ or Fortran). At this stage, we have not yet done this because the R code is more transparent and allows for easy debugging and monitoring but future work will include translating the program into C++ or Fortran.

3. Application to emotional psychophysiology

We will now apply the presented hierarchical state space model to the data that reflect the affective physiological changes in depressed and non-depressed adolescents during conflictual interactions with their parents (Sheeber, Davis, Leve, Hops, & Tildesley, 2007). The study was carried out with adolescents because mood disorders often emerge for the first time during this stage of life (Allen & Sheeber, 2008). Depression, cardiovascular physiology and affective dynamics form a three-node graph for which all three edges have been extensively documented in the literature. For instance, overviews of studies supporting the link between depression and cardiovascular physiology can be found in Carney, Freedland, and Veith (2005), Gupta (2009) and Byrne et al. (in press). Evidence on the link between affective dynamics and depression is documented by Sheeber et al. (2009). The cardiovascular aspects of emotion are described by Bradley and Lang (2007).

In this application section, we will combine the three aforementioned aspects (depression, cardiovascular physiology and affective dynamics) in a state space model to study three specific questions to be outlined below. This section starts with a description of the study and the data, some discussion of the fitted model and basic results and then we consider in detail the three specific research questions.

3.1. Oregon adolescent interaction task

The participants were selected in a double selection procedure in schools, consisting of a screening on depressive symptoms using the Center for Epidemiologic Studies Depression scale (CES-D; Radloff, 1977) followed by diagnostic interviews using the Schedule of Affective Disorders and Schizophrenia-Children's Version (K-SADS; Orvaschel & Puig-Antich, 1994) when the CES-D showed elevated scores. The selected participants were 133 adolescents⁶ (88 females and 45 males) with mean age of 16.2 years. From the total group, 67 adolescents met the DSM-IV criteria for Major Depressive Disorder and were classified as “depressed” whereas 66 adolescents did not meet these criteria and were classified as “normal”.⁷

As part of the study, the adolescent and his/her parents were invited for several nine-minute interaction tasks in the laboratory: two positive tasks (e.g., plan an enjoyable activity), two reminiscence tasks (e.g., discuss salient aspects of adolescence and parenthood) and two conflict tasks (e.g., discuss familial conflicts). During these interactions, second-to-second physiological measures were taken from the adolescent and second-to-second

⁵ The R script can be downloaded at <http://sites.google.com/site/tomlodewyckx/downloads>.

⁶ Originally 141 adolescents participated, but eight participants were excluded from the analysis. For four subjects, the measurement of BP was missing for some or all of the observation moments. For the other four subjects, the time series data were not stationary.

⁷ There were nine out of these 67 depressed individuals who took regular cardiac medication. This fact was ignored in our analysis.

behavioral measures were derived from videotaped behavior of adolescent and parents, resulting in 540 measurements for each task per participant.

The physiological measurements consisted of various measures of heart rate (heart rate, finger pulse heart rate, interbeat interval), blood pressure (blood pressure, diastolic and systolic blood pressure, moving window blood pressure), respiration (respiratory sinus arrhythmia, tidal volume, respiratory period) and skin conductance (skin conductance level, skin conductance response). The behavioral measurements were coded from the videotaped interactions using the Living In Family Environment coding system (LIFE, Hops, Davis, & Longoria, 1995). The behavior of adolescent and parents was classified on a second-to-second basis as either neutral, happy (happy non-verbal or verbal behavior), angry (aggressive/provoking non-verbal or verbal behavior) or dysphoric (sad or anxious non-verbal or verbal behavior).

3.2. Model specification

The Oregon adolescent interaction data set has a large number of variables and a large number of time points. It is impossible to model everything in a single step. Therefore, we have restricted our attention in this paper to a subset of the available data. First, we focus on capturing the dynamics of adolescent's physiology in terms of their heart rate (HR) and blood pressure (BP). Both HR and BP are seen as indicators of sympathetic autonomic arousal and are considered to be part of the emotional stress-response (e.g., Levenson, 1992). Second, in deciding in which context such responses could best be studied, we decided to focus on the conflictual interaction task, as this task was specifically designed to study conflict and the ensuing stress responses. Third, we focused on the dynamics of these physiological indicators and studied the effect of parental anger (denoted as AP). Fourth, we focus only on the third minute (seconds 121–180) of the first conflict task. The reason for the restriction of the number of observation points is that we want to illustrate the usefulness of the hierarchical framework in weakly informative data. Thus, to summarize, the three time-varying measures of interest are HR, BP (both outputs) and AP (input), and this in the third minute of the first conflictual task.

HR was derived from the measured electrocardiograph (ECG) signal and its unit is BPM (beats per minute). BP was obtained with the Portapres portable device, with its unit being mmHg (millimeter of mercury). The behavioral measure AP was obtained by aggregation of two measures that were obtained after analysis of the videotaped interactions with the LIFE coding system. The constituent variables AngerFather (AF) and AngerMother (AM) were binary: They were equal to 0 when the behavior of respectively father or mother was classified an “not angry” and 1 if it was classified as “angry”. Then AP was derived as the sum of AF and AM, taking values of 0 (none of the parents shows angry behavior), 1 (at least one parent shows angry behavior) and 2 (both parents show angry behavior). In Fig. 2, an example of the data is shown for two random participants.

The hierarchical state space model is based on the observable outputs HR and BP and the observable input (or state covariate) AP. The model for individual i is

$$\mathbf{y}_{t,i} | \boldsymbol{\theta}_{t,i} \sim N(\boldsymbol{\mu}_i + \boldsymbol{\theta}_{t,i}, \boldsymbol{\Sigma}_\epsilon)$$

$$\boldsymbol{\theta}_{t,i} | \boldsymbol{\theta}_{t-1,i}, \dots, \boldsymbol{\theta}_{t-5,i} \sim N\left(\sum_{l=1}^5 \boldsymbol{\Phi}_{l,i} \boldsymbol{\theta}_{t-l,i} + \boldsymbol{\Delta}_{i,t} \mathbf{z}_{t-1,i}, \boldsymbol{\Sigma}_\eta\right) \quad (14)$$

where the vector of observed processes $\mathbf{y}_{t,i}$ consists of $y_{[1],t,i}$ (observed HR of individual i at time t) and $y_{[2],t,i}$ (observed BP of individual i at time t), the vector of latent processes $\boldsymbol{\theta}_{t,i}$ consists of $\theta_{[1],t,i}$ (latent HR of individual i at time t) and $\theta_{[2],t,i}$ (latent BP of individual i at time t) and the scalar covariate $\mathbf{z}_{t-1,i}$ equals the

observation of AP at time $t - 1$ for individual i (we were interested in the effect of parental anger on the adolescent's physiology after one second). The number of lags for the transition matrices was set at 5 s (lag 5), which is consistent with earlier findings that this is the appropriate time window to look at (see e.g., Matsukawa & Wada, 1997).

The random effects distributions for the means, covariate effects and transition matrices are defined as in Eqs. (8), (11) and (12) respectively. The mean vector $\boldsymbol{\alpha}_\mu$ consists of the elements α_{μ_1} (hierarchical mean of HR) and α_{μ_2} (hierarchical mean of BP), and the diagonal of \mathbf{B}_μ consists of the elements β_{μ_1} and β_{μ_2} (the corresponding hierarchical variances). The population mean $\boldsymbol{\alpha}_{\phi_l}$ has elements $\alpha_{\phi_{[1],1|l}}$ (hierarchical mean of the autoregressive effect HR for lag l), $\alpha_{\phi_{[1],2|l}}$ (hierarchical mean of the crosslagged effect BP on HR for lag l), $\alpha_{\phi_{[2],1|l}}$ (hierarchical mean of the crosslagged effect HR on BP for lag l) and $\alpha_{\phi_{[2],2|l}}$ (hierarchical mean of the autoregressive effect BP for lag l). In addition, the diagonal of \mathbf{B}_{ϕ_l} contains elements $\beta_{\phi_{[1],1|l}}$, $\beta_{\phi_{[1],2|l}}$, $\beta_{\phi_{[2],1|l}}$ and $\beta_{\phi_{[2],2|l}}$ (the corresponding hierarchical variances). Finally, $\boldsymbol{\alpha}_\Delta$ consists of the elements α_{δ_1} (hierarchical mean of the regression effect of AP on HR) and α_{δ_2} (hierarchical mean of the regression effect of AP on BP) and the diagonal of \mathbf{B}_Δ consisting of the elements β_{δ_1} and β_{δ_2} (the corresponding hierarchical variances).

The introduced hierarchical state space model is estimated separately for data of the normal and the depressed adolescents, using a Gibbs sampler that was implemented in R. The prior distributions were chosen to be conjugate to the model likelihood.⁸ For each analysis, we obtained three independent chains of 6000 samples of which the first 1000 were discarded as burn-in. To monitor convergence of the MCMC chains, \hat{R} (Gelman et al., 2004) was calculated for each single parameter and convergence was obtained for most parameters according to \hat{R} .⁹ In Table 1, the posterior medians and the 95% credibility intervals are reported for selected parameters from the analysis of the normal and depressed adolescents. This is to give the reader a general idea of the values of the parameters in the estimated model. Specific parameters will be discussed below in the context of specific questions. We will elaborate on this in the following paragraphs.

3.3. Question 1: Tachycardia & hypertension?

The literature on psychophysiology and psychiatry recognizes that Major Depressive Disorder is associated with various physiological processes (Gupta, 2009). For instance, tachycardia (elevated heart rate) was found to be linked to depression (Byrne et al., in press; Carney et al., 2005; Dawson, Schell, & Catania, 1977; Lahmeyer & Bellur, 1987). More extensively studied is the link between depression and hypertension (high blood pressure): Depressed individuals seem to generally have a higher blood pressure than normals (Davidson, Jonas, Dixon, & Markovitz, 2000; Jonas, Franks, & Ingram, 1997; Rutledge & Hogan, 2002; Scherrer et al., 2003), although in some studies the effect was not replicated (Lake et al., 1982; Wiehe et al., 2006) or even a reverse effect

⁸ For the diagonal elements of the covariance matrices $\boldsymbol{\Sigma}_\eta$, $\boldsymbol{\Sigma}_\epsilon$, \mathbf{B}_μ , \mathbf{B}_{ϕ_l} and \mathbf{B}_Δ , we choose scaled inverse χ^2 priors with degrees of freedom $\nu_0 = 1$ and scale $s_0 = 1$. Although these are informative priors, the influence of the prior distribution was negligible compared to the overwhelming influence of the data. For the hierarchical mean vectors $\boldsymbol{\alpha}_\mu$, $\boldsymbol{\alpha}_\phi$ and $\boldsymbol{\alpha}_\Delta$, uninformative multivariate normal priors are assumed with a mean vector $\boldsymbol{\mu}_0 = \mathbf{0}$ being a vector of zeros, a covariance matrix $\boldsymbol{\Sigma}_0 = 10^6 \times \mathbf{I}$ being the identity matrix with the value of 10^6 on the diagonal and $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ having the appropriate dimensions. For more details, see Appendix A.

⁹ $\hat{R} < 1.10$ for all parameters and both populations, except for four elements of $\boldsymbol{\Sigma}_\epsilon$ and \mathbf{B}_Δ as a consequence of strongly autocorrelated Markov chains (the \hat{R} values are 1.16, 1.43, 1.56 and 2.49). This problem can be solved by increasing sample size and thinning the chains.

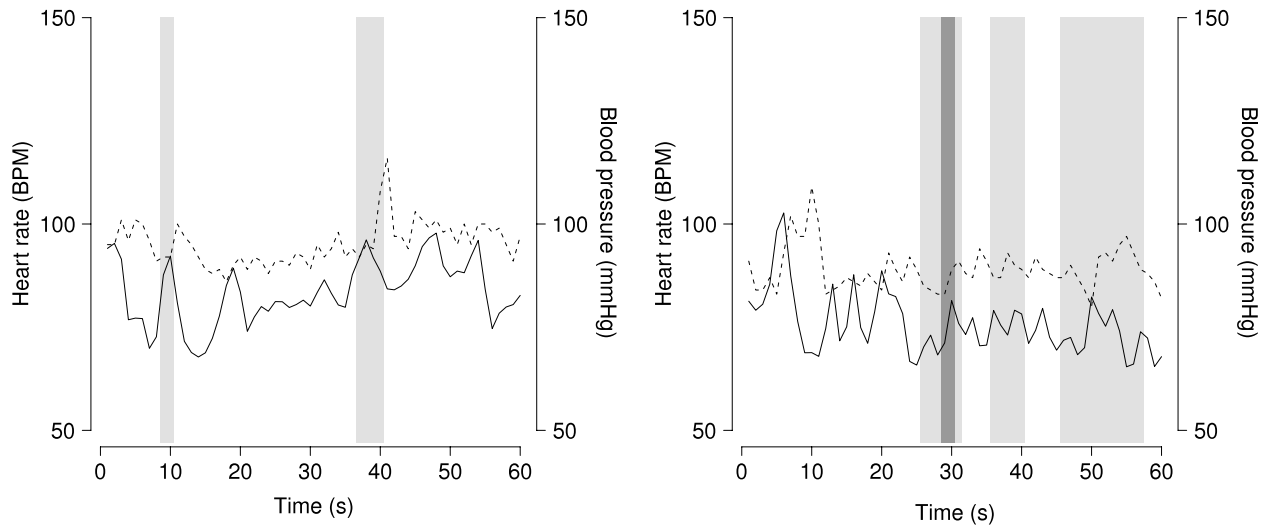


Fig. 2. Example data for two randomly chosen participants, showing their fluctuations in heart rate (HR; full line), blood pressure (BP; dotted line) and anger parents (AP; background shading: white = 0, light grey = 1 and dark grey = 2).

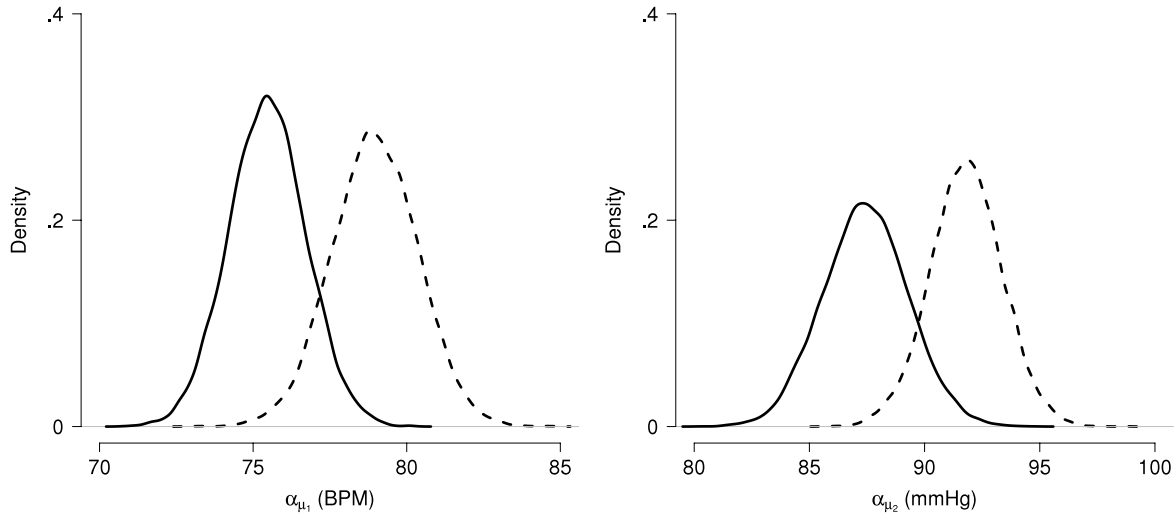


Fig. 3. Estimated posterior densities for normals (full lines) and depressed (dotted lines) of the parameters α_{μ_1} and α_{μ_2} , the hierarchical means of respectively HR and BP.

was observed (Licht et al., 2009). Various explanations for the link between hypertension and depression were investigated such as psychosocial stressors (Bosworth, Bartash, Olsen, & Steffens, 2003; Sparrenberger et al., 2006), antidepressant use (Licht et al., 2009) and sleep (Gangwisch et al., 2010).

To investigate this particular difference in physiology for depressed and non-depressed individuals, we focus in the hierarchical state space model on the estimated posterior distributions for the elements of the hierarchical parameter vector α_{μ} for the groups of normals and depressed individuals. In Fig. 3, the estimated posterior densities for the elements α_{μ_1} and α_{μ_2} are compared and visual inspection suggests that average heart rate and blood pressure are higher for depressed than for normals.

As a formal test, Bayes factors are estimated to quantify the support of the comparisons. Bayes factor estimation is performed with the encompassing prior method (see e.g., Hoijtink, Klugkist, & Boelen, 2008; Klugkist & Hoijtink, 2007). This method easily allows to estimate the Bayes factor to test whether a parameter θ is larger or smaller than a fixed value. Given that the prior distribution of θ gives equal prior mass to both scenarios $M_1 : \theta < 0$ and $M_2 : \theta > 0$, an estimate for the Bayes factor BF_{21} (in favor of M_2) is equal to the ratio of the posterior proportions of parameter θ being consistent with each of the models: $\hat{BF}_{21} = \hat{Pr}(\theta > 0 | Y) / \hat{Pr}(\theta < 0 | Y)$. To test a difference between the populations of normal and depressed individuals, a difference parameter was

derived (e.g., $\theta^{\text{diff}} = \theta^{\text{depr}} - \theta^{\text{norm}}$) and the encompassing prior method was applied to the estimated posterior distribution of this difference parameter.

We find strong support for heart rate ($BF = 30.12$) and blood pressure ($BF = 25.60$) being higher for depressed than for normals.¹⁰ These findings are consistent with many studies in the literature that suggest a link between depression, tachycardia and hypertension. Although the particular setting of our study strongly differs from most studies about tachycardia and hypertension, our results demonstrate that the associations with depression remain even during stressful interactions.

3.4. Question 2: Emotional inertia?

As mentioned in the introduction, emotions are dynamic in nature. Not “despite” but “because of” the fact that our emotional state constantly changes, we are able to experience emotions. The concept of emotional inertia poses that an individual

¹⁰ For the interpretation of the values of the Bayes factor estimates, we use the interpretation scheme by Raftery (1995): $BF_{21} > 150 \rightarrow$ Very strong support M_2 ; $BF_{21} \in [20, 150] \rightarrow$ Strong support M_2 ; $BF_{21} \in [3, 20] \rightarrow$ Positive support M_2 ; $BF_{21} \in [1, 3] \rightarrow$ Weak support M_2 ; $BF_{21} = 1 \rightarrow$ No support for either model; $BF_{21} \in [0, 1] \rightarrow$ Support M_1 , strength of support for M_1 is evaluated by taking the inverse value and interpreting it parallel to the aforementioned categories.

Table 1

Posterior medians ($q_{0.500}$) and 95% credibility interval ($[q_{0.025}, q_{0.975}]$) for a selection of parameters in the analysis of the hierarchical model for the normal and depressed adolescents. Abbreviations used are HM (hierarchical mean), HV (hierarchical variance), AR (autoregression), CR (crossregression), HR (heart rate), BP (blood pressure) and AP (anger parents).

	Description	Normals			Depressed		
		$q_{0.025}$	$q_{0.500}$	$q_{0.975}$	$q_{0.025}$	$q_{0.500}$	$q_{0.975}$
<i>Hierarchical parameters μ</i>							
α_{μ_1}	HM for general mean HR	73.01	75.46	77.92	76.19	78.96	81.69
α_{μ_2}	HM for general mean BP	83.86	87.44	91.10	88.67	91.82	94.88
β_{μ_1}	HV for general mean HR	70.98	99.94	144.43	91.33	127.71	184.43
β_{μ_2}	HV for general mean BP	157.05	221.30	320.31	113.90	158.09	229.16
<i>Hierarchical parameters Φ_1</i>							
$\alpha_{\phi_{1,1 1}}$	HM for AR effect HR	0.85	0.93	1.01	0.85	0.93	1.00
$\alpha_{\phi_{1,2 1}}$	HM for CR effect BP \rightarrow HR	−0.23	−0.16	−0.10	−0.22	−0.17	−0.11
$\alpha_{\phi_{2,1 1}}$	HM for CR effect HR \rightarrow BP	0.11	0.17	0.22	0.11	0.17	0.23
$\alpha_{\phi_{2,2 1}}$	HM for AR effect BP	0.40	0.46	0.53	0.33	0.40	0.47
$\beta_{\phi_{1,1 1}}$	HV for AR effect HR	0.04	0.06	0.09	0.03	0.04	0.07
$\beta_{\phi_{1,2 1}}$	HV for CR effect BP \rightarrow HR	0.03	0.04	0.07	0.02	0.04	0.05
$\beta_{\phi_{2,1 1}}$	HV for CR effect HR \rightarrow BP	0.02	0.03	0.05	0.02	0.04	0.06
$\beta_{\phi_{2,2 1}}$	HV for AR effect BP	0.04	0.06	0.08	0.04	0.06	0.09
<i>Hierarchical parameters Δ</i>							
α_{δ_1}	HM for effect AP \rightarrow HR	−0.79	−0.27	0.23	−0.88	−0.38	0.13
α_{δ_2}	HM for effect AP \rightarrow BP	0.14	0.78	1.47	−0.12	0.38	0.86
β_{δ_1}	HV for effect AP \rightarrow HR	0.13	0.41	1.60	0.30	1.08	3.49
β_{δ_2}	HV for effect AP \rightarrow BP	0.27	1.26	4.17	0.13	0.41	1.26
<i>Error variances</i>							
$\sigma_{\epsilon,1}^2$	Measurement error variance HR	0.25	0.58	1.03	0.49	0.82	1.30
$\sigma_{\epsilon,2}^2$	Measurement error variance BP	0.22	0.79	1.51	0.12	0.40	0.99
$\sigma_{\eta,1}^2$	Innovation variance HR	16.66	17.59	18.60	14.63	15.47	16.39
$\sigma_{\eta,2}^2$	Innovation variance BP	15.50	16.34	17.2	21.57	22.65	23.78

lingers in a certain emotional state for a while (Kuppens, Allen, & Sheeber, 2010). A high level of emotional inertia implies a lack of flexibility in adapting emotions to the surrounding context and in response to regulation efforts. In line with this reasoning, Kuppens et al. (2010) demonstrated that the emotional behavior or experience of depressed individuals is characterized by higher levels of inertia than that of their non-depressed counterparts. An important research question in this respect is whether this increased emotional inertia observed in the behavioral expression of emotions of depressed individuals also extends to the dynamics of physiological emotion components. Along these lines, previous research has suggested that depressed individuals are characterized by slower HR recovery after physical exercise (Hughes et al., 2006). The question is, however, whether such slower recovery will also be observed in a psychologically demanding setting, a finding that would moreover be in line with the idea of higher inertia.

To explore the dynamics of state space models, we can work with impulse response functions (see Hamilton, 1994) which have been applied extensively in the domain of cardiovascular modeling (see e.g., Matsukawa & Wada, 1997; Nishiyama, Yana, Mizuta, & Ono, 2007; Panerai, James, & Potter, 1997; Triedman, Perrott, Cohen, & Saul, 1995). The impulse response function represents how a one unit innovation impulse of a process affects the process itself (autoregressive impulse response) or another process (crossregressive impulse response) over time. This temporal effect is compared to a baseline of 0, to be interpreted as “there was no impulse”. The values in the impulse response functions are calculated from parameter estimates in time series models (for more details about the calculation, see Hamilton, 1994; Matsukawa & Wada, 1997).

In Fig. 4, the autoregressive and crossregressive impulse response functions for HR and BP are shown for the groups of normal and depressed adolescents. These response functions were derived from the posterior means of α_ϕ and reflect the impulse response functions at the hierarchical level. To guide interpretation, let us

first focus on the autoregressive effect for HR in Fig. 4(a). When HR is incremented with a “pulse” of one BPM at time $t = 0$, it stays elevated for $t = 1, \dots, 4$. However, around 5 s, a downward correction occurs (until about 8 s) and thereafter, the effect fades out. The autoregressive impulse response function for BP in Fig. 4(d) shows a recovery to “no effect” in about 5 s and has a less obvious oscillatory component. These impulse effects are similar for the normal and depressed population, which suggest that emotional inertia for physiological emotion components is not stronger for depressed individuals, contrary to what was found for the behavioral expression of emotion (Kuppens et al., 2010), which displayed increased inertia in depression.¹¹ These results thus point to a possible dissociation between the dynamical properties of emotional expression and physiology, indicating that both may be governed by different types of regulatory control. Expressive emotional behavior can indeed be thought to be strongly governed by social and self-representational concerns (like display rules, see Darwin, Ekman, & Prodger, 2002), whereas physiology is obviously primarily dictated by much more autonomous processes. The fact that higher levels of inertia in depression are observed for the former rather than the latter processes, may thus indicate that the emotional dysfunctioning associated with this mood disorder may be particularly manifested in emotion components that have implications for the social world (for the importance of social concerns in depression, see e.g., Allen & Badcock, 2003).

When considering the crossregressive impulse response functions in Fig. 4(b) and (c), we find them to be strongly asymmetric. This observation is consistent with previous studies and reflects a basic dynamical relation of these cardiovascular processes (Matsukawa & Wada, 1997; Nishiyama et al., 2007; Triedman et al., 1995).

¹¹ We could not use Bayes factors to compare the IRFs, so we relied upon visual inspection here.

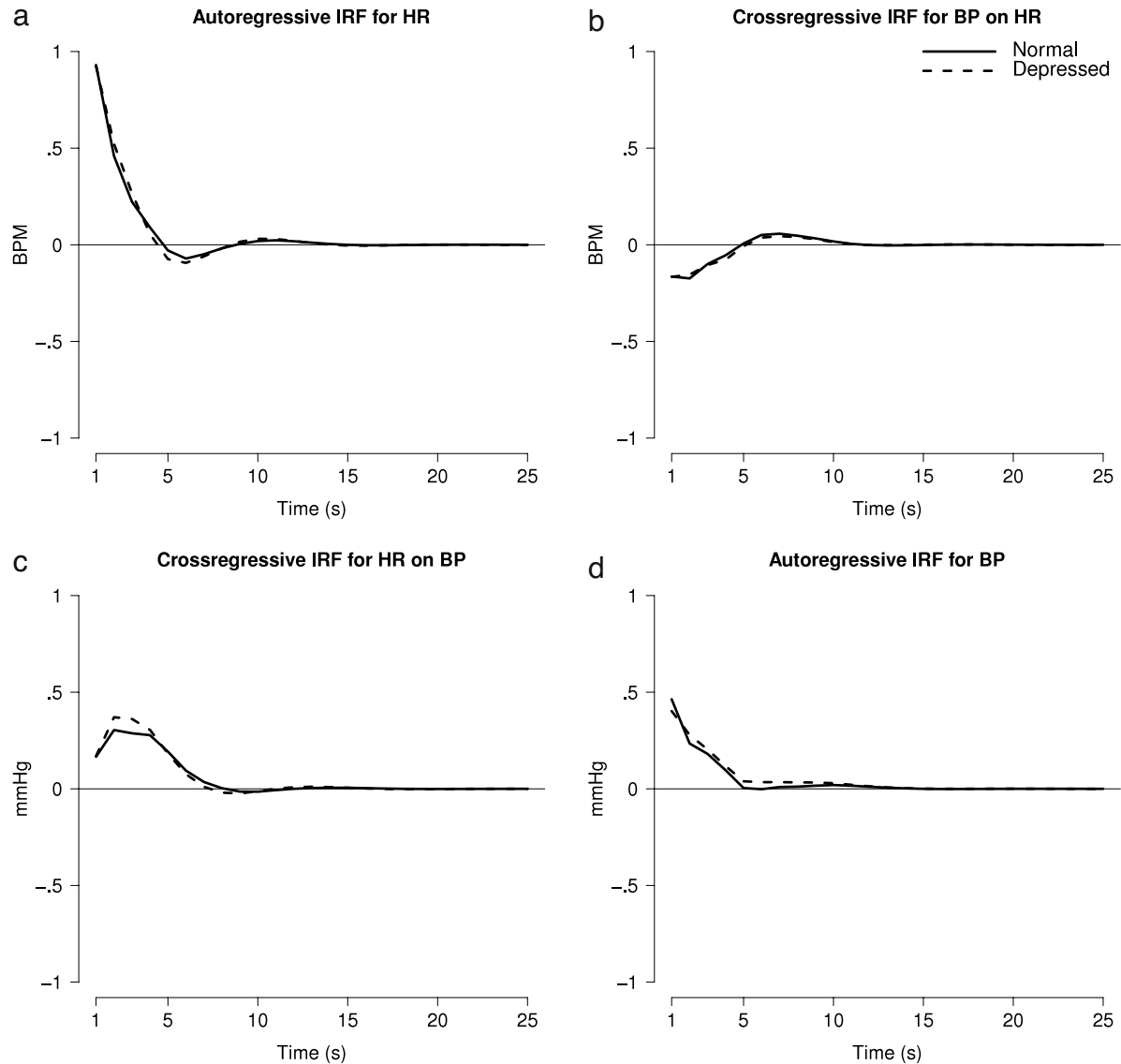


Fig. 4. Hierarchical impulse response functions for the groups of normal and depressed individuals with (a) the HR autoregression effect, (b) the BP to HR crossregression effect, (c) the HR to BP crossregression effect and (d) the BP autoregression effect. The impulse response functions were derived from the posterior median of the hierarchical mean vector α_Φ , as explained in Hamilton (1994).

The impulse response functions in Fig. 4 handled the dynamical patterns of the physiological processes on population level. However, impulse response functions can also be derived for each individual i , based on estimated posterior medians of $[\Phi_{i,1}, \dots, \Phi_{i,5}]$. In Fig. 5, the individual autoregressive impulse response functions are shown for 10 random participants of both groups, with the estimated population impulse response functions as a reference. This illustrates the individual differences within the populations. The strong variations reflect the problem of data: Only very few data are available for each individual, resulting in quite unique forms for the individual impulse response functions. The hierarchical parameter distribution stabilizes these individual estimates and allows for inference at the population level. Moreover, the complex forms of the impulse response functions rationalize our approach to choose a lag equal to 5. If only one lag was relevant, all impulse response functions would fade out monotonously without any quadratic or sinusoidal tendency.

3.5. Question 3: Emotional reactivity?

The research literature on emotional reactivity in depression argues that depressed differ from non-depressed individuals in how they emotionally respond to specific eliciting events or stimuli. The findings regarding the nature or direction of this difference are inconsistent, however. In terms of reactivity to negative stimuli or events, on the one hand, there is research suggesting that depressed individuals are characterized by increased reactivity to negative or aversive stimuli, while on the other hand there is also research that points to the opposite, namely that depressed are characterized by decreased emotional reactivity, something that has been labeled emotion context insensitivity (ECI; Rottenberg, 2005). A recent meta-analysis of laboratory studies found most support for the ECI view across self-report, behavioral and physiological systems (Bylsma, Morris, & Rottenberg, 2008), but there remained a great deal of variability between findings from different studies. However, most of this research examined emotional

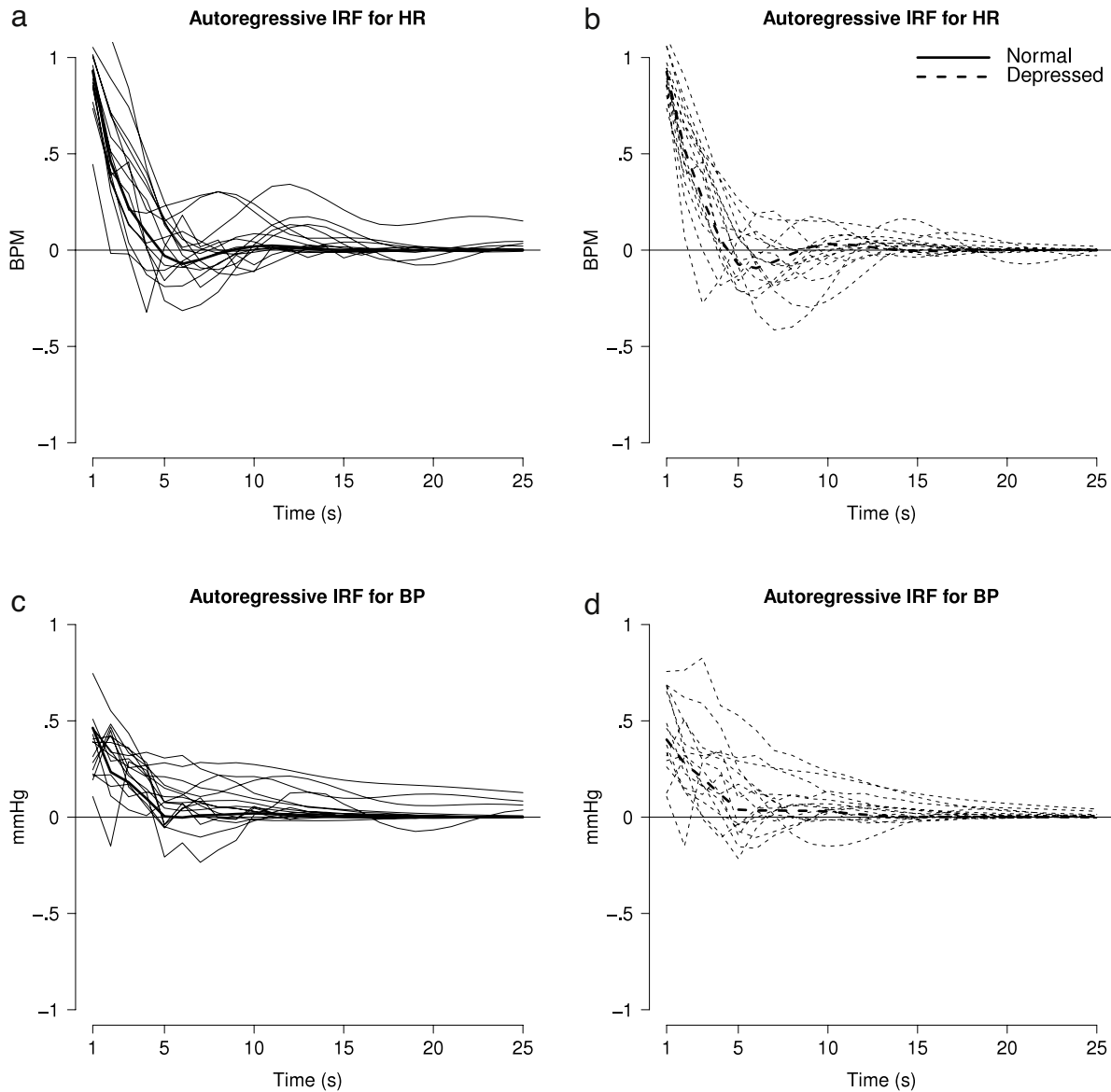


Fig. 5. Individual differences in the autoregressive IRF for HR in normals and depressed, respectively (a) and (b), and individual differences in the autoregressive IRF for BP in normals and depressed, respectively (c) and (d). The bold lines represent the hierarchical IRF, the thin lines the individual ones. Each plot is based on a random subset of 15 participants. The impulse response functions were derived as explained in Hamilton (1994), with the individual functions based on the posterior medians of the set of transition matrices $[\Phi_{1,i}, \Phi_{2,i}, \Phi_{3,i}, \Phi_{4,i}, \Phi_{5,i}]$ and the hierarchical functions based on the posterior median of the hierarchical vector α_Φ .

responding to standardized, non-idiographic stimuli (such as standardized emotional pictures or film clips). In contrast, the current data allows us to study emotional reactivity in a context that is highly personally relevant to the individual (i.e., their own family environment). The question that arises is thus whether depressed will indeed differ from non-depressed in how they emotionally respond to a personally meaningful and stressful event (anger displayed by the parent), and if so, in what way they differ, as indicated by changes in their HR and BP.

The parameter of interest here is α_Δ , the hierarchical mean vector of the regression coefficient vectors δ_i , containing the effects of AP. In Fig. 6, the estimated posterior densities of the elements α_{δ_1} and α_{δ_2} are shown for the depressed and normal groups. Concerning HR (see left panel of Fig. 6), we found positive support in favor of an inhibitory effect of parental anger for both depressed ($\hat{BF} = 0.08$) and normals ($\hat{BF} = 0.17$), and these effects are found to be equally strong ($\hat{BF} = 0.62$). A heart rate deceleration has been theoretically linked to an orienting response, related

to information intake and orientation towards a salient stimulus (Cook & Turpin, 1997). The observed HR response thus suggests that participants are alerted by anger expression by the parent and orient their attention to such behavior.

On the other hand, concerning BP (see right panel of Fig. 6), positive ($\hat{BF} = 11.69$) and strong ($\hat{BF} = 89.36$) support was found in favor of an excitatory effect of parental anger for respectively depressed and normal individuals. The effect is stronger for normal individuals as compared to depressed ($\hat{BF} = 0.19$). These results show that angry parental behavior elicits a blood pressure increase, indicative of a stress response. The fact that the effect is much stronger for normals than for depressed is consistent with the ECI hypothesis, in that depressed participants do not show the typical emotional reactivity observed in normal individuals, even in response to a stressful, idiographically highly relevant event (parental anger). Instead, they are characterized by a relative insensitivity to such stimuli, in line with predictions and findings from emotion context sensitivity research (Rottenberg, 2005).

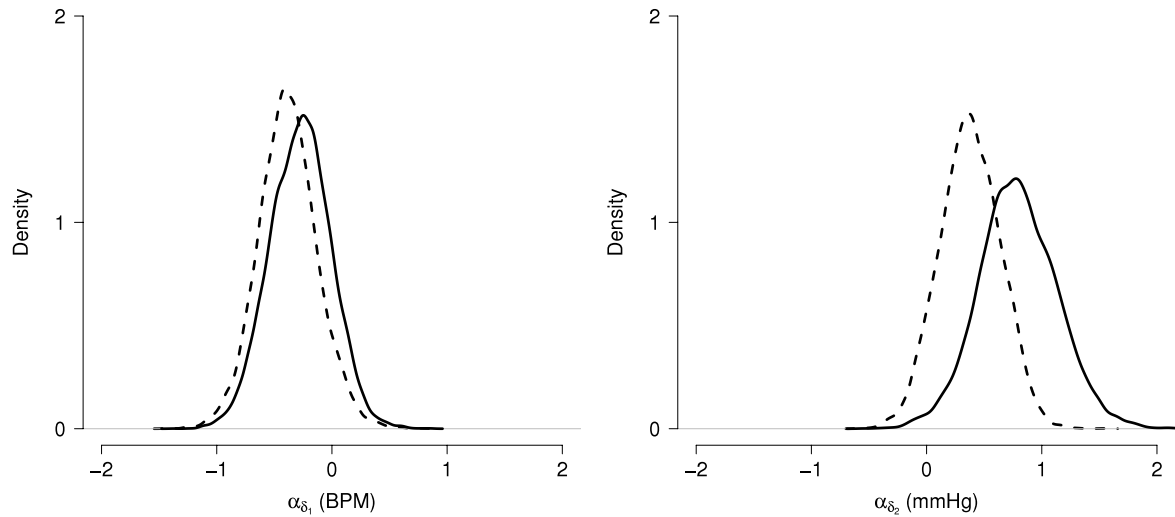


Fig. 6. Estimated posterior densities for normals (full lines) and depressed (dotted lines) of the parameters α_{δ_1} and α_{δ_2} , the hierarchical regression effects of AP on respectively HR and BP.

4. Discussion

In this paper, we introduced a hierarchical extension for the linear Gaussian state space model. The extended model proved to be very useful when studying individual differences in the dynamics of emotional systems. Applying it to the Oregon adolescent interaction data led to interesting discussions about hypotheses on the relations between cardiovascular processes, emotion dynamics and depression. On the whole, it was clear that applying the hierarchical state space approach to these data provides a detailed picture of several central characteristics of the dynamics underlying physiological emotion components and its determinants (both contextual and individual differences). As such, it allowed us to address several research questions that are central to the study of emotion dynamics in relation to psychopathology.

Although the model is very flexible, we must remark that it has several strong assumptions that we might want to relax in the future. For instance, the assumptions of linearity and Gaussian error distributions might not always be realistic, especially when considering psychological data (e.g., outliers, sudden changes in level, discrete processes). Also, random effects distributions for the error covariance matrices might be added, allowing to study differences in for instance the link between heart rate variability and depression (Frasure-Smith, Lesperance, Irwin, Talajic, & Pollock, 2009; Licht et al., 2008; Pizzi, Manzoli, Mancini, & Costa, 2008). However, for some generalizations, the appealing traditional Gibbs sampling approach with closed form full conditionals, as discussed in Appendix A, might not be possible anymore.

Regarding more significant extensions for the future, an explanatory model can be added on the hierarchical level, such that individual differences can be interpreted in function of other individual characteristics (e.g., emotional inertia is associated to only particular depression symptoms). Further, the design matrix Ψ (which was fixed to an identity matrix in our approach) can be considered as a free parameter for each individual, which turns the model into a hierarchical dynamic factor model. Finally, complex hypotheses might be tested when discrete output variables might be allowed, whereas the truly observed discrete outputs are connected to the underlying unobserved continuous outputs through linking functions.

Maybe one of the most important advantages of the model is its generality: In practice, it can be applied beyond the context of physiological emotion responses, to any type of time series and in various research domains, whether one wants to make inferences about the values of the latent states or investigate differences in the system parameters for several dynamical systems.

Appendix A. Derivation of full conditionals distributions

A Gibbs sampler consists of full conditionals for single parameters or blocks of parameters. When iteratively sampling from these full conditional distributions, convergence is obtained in infinity.¹² After convergence, samples are considered to be simulations of the true posterior distribution of these parameters and thus these posterior distributions can be approximated with a large set of samples. The main principle of deriving these full conditionals is to combine the likelihood function of the parameter with its prior distribution. To obtain a closed form distribution, the likelihood and prior should match, i.e., some kernel elements in their formulas should be combined such that a new distribution is obtained. As the likelihood function is determined by the model, which is fixed, the prior distribution can be chosen in such a way that it can be combined with the likelihood, which is called a conjugate prior.

In the section on statistical inference, we distinguished three categories of parameters in the Gibbs sampler: (1) the latent states, (2) the system parameters and (3) the hierarchical parameters. The latent states were sampled with the forward filtering backward sampling algorithm (Carter & Kohn, 1994; Kim & Nelson, 1999), as explained in that section, and the remaining parameters are sampled using full conditional distributions that are derived by combining the likelihood function with conjugate prior distributions. In this appendix, we provide more details on the derivation of these full conditionals. For every type of parameter, we illustrate the technique for one parameter and generalize it to the other parameters of the same type. The parameter types that are discussed in the respective order are variances, hierarchical mean vectors and individual effects.

A.1. Full conditionals for variances

The covariance matrices in the hierarchical state space model, as described in Eqs. (7), (8) and (10)–(12), are Σ_ϵ , Σ_η , B_μ , B_ϕ , B_Γ and B_Δ . All matrices are assumed to be diagonal matrices, with the variances on the diagonal and the covariances on the off-diagonal elements fixed to 0.¹³ We illustrate how to derive the full conditional for β_{μ_1} , i.e., the first diagonal element of B_μ .

¹² In theory, the sample size should go to infinity. In practice, convergence can be evaluated with convergence statistics, like \hat{R} (Gelman et al., 2004).

¹³ Another approach is to consider the covariance matrix as a whole and in such a case, the conjugate inverse Wishart prior distribution leads to an inverse Wishart full conditional. This will not be discussed here.

In the hierarchical state space model, the only equation that is relevant for the likelihood function of \mathbf{B}_μ is Eq. (8). As all covariances in \mathbf{B}_μ are fixed to 0 and thus all elements in vector μ_i are independent, we can isolate the first element $\mu_{[1]i}$ and express it as a univariate Normal,

$$\mu_{[1]i} \sim N(\alpha_{\mu_1}, \beta_{\mu_1}).$$

The likelihood function for β_{μ_1} can be expressed and simplified as follows:

$$\begin{aligned} \ell(\beta_{\mu_1}) &= \prod_{i=1}^I N(\mu_{[1]i} | \alpha_{\mu_1}, \beta_{\mu_1}) \\ &\propto \prod_{i=1}^I (\beta_{\mu_1})^{-\frac{1}{2}} \exp\left(-\frac{(\mu_{[1]i} - \alpha_{\mu_1})^2}{2\beta_{\mu_1}}\right) \\ &\propto (\beta_{\mu_1})^{-\frac{I}{2}} \exp\left(-\frac{\sum_{i=1}^I (\mu_{[1]i} - \alpha_{\mu_1})^2}{2\beta_{\mu_1}}\right) \\ &\propto (\beta_{\mu_1})^{-\frac{I}{2}} \exp\left(-\frac{SS}{2\beta_{\mu_1}}\right), \end{aligned} \quad (15)$$

with SS being the sum of squared residuals $\sum_{i=1}^I (\mu_{[1]i} - \alpha_{\mu_1})^2$. In this structure, the kernel of a scaled inverse χ^2 distribution is recognized. Therefore, a conjugate scaled inverse χ^2 prior distribution for β_{μ_1} is chosen, with ν_0 degrees of freedom and scale s_0^2 , resulting in a scaled inverse χ^2 full conditional distribution with ν_1 degrees of freedom and scale s_1^2 :

$$\begin{aligned} P(\beta_{\mu_1} | \mathbf{D}, \mathbf{P}) &\propto P(\beta_{\mu_1}) \ell(\beta_{\mu_1}) \\ &\propto (\beta_{\mu_1})^{-\frac{\nu_0}{2}+1} \exp\left(-\frac{\nu_0 s_0^2}{2\beta_{\mu_1}}\right) (\beta_{\mu_1})^{-\frac{I}{2}} \exp\left(-\frac{SS}{2\beta_{\mu_1}}\right) \\ &\propto (\beta_{\mu_1})^{-\frac{\nu_0+I}{2}+1} \exp\left(-\frac{\nu_0 s_0^2 + SS}{2\beta_{\mu_1}}\right) \\ &\sim SI\chi^2(\nu_1, s_1^2), \end{aligned} \quad (16)$$

with \mathbf{D} representing the set of all data and \mathbf{P} representing the set of all model parameters but β_{μ_1} . The specific values of the full conditional parameters are

$$\begin{aligned} \nu_1 &= \nu_0 + I \\ s_1^2 &= \frac{\nu_0 s_0^2 + SS}{\nu_1}. \end{aligned}$$

The derivation is parallel for all other variances. To obtain ν_1 , the prior degrees of freedom ν_0 are incremented with the “amount of residuals”, which is I for the random effects variances on the diagonals of \mathbf{B}_μ , \mathbf{B}_ϕ , \mathbf{B}_Γ and \mathbf{B}_Δ and IT for the error variances on the diagonals of Σ_ϵ and Σ_η (these parameters are constant over I individuals and T measurement occasions). To obtain s_1^2 , we need to calculate SS , the sum of squared residuals. For the random effects variances, the residuals are the differences between the individual parameters and hierarchical means for $i = 1, \dots, I$. For the error variances in Σ_ϵ and Σ_η , the residuals are the deviations of the data or latent states to the mean term in respectively the observation and transition equation. For instance, the residuals in the observation equation are equal to $\epsilon_{t,i} = \mathbf{y}_{t,i} - \mu_i - \theta_{t,i} - \Gamma_i \mathbf{x}_{t,i}$ for $t = 1, \dots, T$ and $i = 1, \dots, I$. When focusing on a particular variance on the diagonal of Σ_ϵ , the corresponding vector element in $\epsilon_{t,i}$ should be considered.

A.2. Full conditionals for hierarchical mean vectors

The hierarchical mean vectors in the hierarchical state space model are α_μ , α_ϕ , α_Γ and α_Δ . Contrary to the covariance matrices,

these vector elements are not independent in the sampling process. We illustrate the derivation for α_μ .

The only equation in the hierarchical state space model that contains α_μ is Eq. (8) (repeated):

$$\mu_i \sim N(\alpha_\mu, \mathbf{B}_\mu),$$

thus, the likelihood for α_μ is expressed as

$$\begin{aligned} \ell(\alpha_\mu) &= \prod_{i=1}^I N(\mu_i | \alpha_\mu, \mathbf{B}_\mu) \\ &\propto \prod_{i=1}^I \exp\left(-\frac{1}{2} [(\mu_i - \alpha_\mu)' \mathbf{B}_\mu^{-1} (\mu_i - \alpha_\mu)]\right) \\ &\propto \exp\left(-\frac{1}{2} \left[\sum_{i=1}^I (\alpha_\mu' \mathbf{B}_\mu^{-1} \alpha_\mu) \right. \right. \\ &\quad \left. \left. - \sum_{i=1}^I (\mu_i' \mathbf{B}_\mu^{-1} \alpha_\mu) - \sum_{i=1}^I (\alpha_\mu' \mathbf{B}_\mu^{-1} \mu_i) \right] \right) \\ &\propto \exp\left(-\frac{1}{2} [(\alpha_\mu' [\mathbf{I} \mathbf{B}_\mu^{-1}] \alpha_\mu) - (\mathbf{S}_\mu' \mathbf{B}_\mu^{-1} \alpha_\mu) - (\alpha_\mu' \mathbf{B}_\mu^{-1} \mathbf{S}_\mu)]\right), \end{aligned} \quad (17)$$

where $\mathbf{S}_\mu = \sum_{i=1}^I \mu_i$. Now, a conjugate multivariate normal prior for α_μ is formulated as

$$\alpha_\mu \sim N(\mu_0, \Sigma_0).$$

The result is a multivariate normal full conditional distribution for α_μ :

$$\begin{aligned} P(\alpha_\mu | \mathbf{D}, \mathbf{P}) &\propto P(\alpha_\mu) \ell(\alpha_\mu) \\ &\propto \exp\left(-\frac{1}{2} [(\alpha_\mu - \mu_0)' \Sigma_0^{-1} (\alpha_\mu - \mu_0)]\right) \\ &\quad \times \exp\left(-\frac{1}{2} [(\alpha_\mu' [\mathbf{I} \mathbf{B}_\mu^{-1}] \alpha_\mu) - (\mathbf{S}_\mu' \mathbf{B}_\mu^{-1} \alpha_\mu) - (\alpha_\mu' \mathbf{B}_\mu^{-1} \mathbf{S}_\mu)]\right) \\ &\propto \exp\left(-\frac{1}{2} [(\alpha_\mu' \Sigma_0^{-1} \alpha_\mu) - (\mu_0' \Sigma_0^{-1} \alpha_\mu) - (\alpha_\mu' \Sigma_0^{-1} \mu_0)]\right) \\ &\quad \times \exp\left(-\frac{1}{2} [(\alpha_\mu' [\mathbf{I} \mathbf{B}_\mu^{-1}] \alpha_\mu) - (\mathbf{S}_\mu' \mathbf{B}_\mu^{-1} \alpha_\mu) - (\alpha_\mu' \mathbf{B}_\mu^{-1} \mathbf{S}_\mu)]\right) \\ &\propto \exp\left(-\frac{1}{2} [\alpha_\mu' (\Sigma_0^{-1} + \mathbf{I} \mathbf{B}_\mu^{-1}) \alpha_\mu \right. \\ &\quad \left. - (\mu_0' \Sigma_0^{-1} + \mathbf{S}_\mu' \mathbf{B}_\mu^{-1}) \alpha_\mu - \alpha_\mu' (\Sigma_0^{-1} \mu_0 + \mathbf{B}_\mu^{-1} \mathbf{S}_\mu)]\right) \\ &\sim N(\mu_1, \Sigma_1) \end{aligned} \quad (18)$$

where

$$\begin{aligned} \Sigma_1 &= (\Sigma_0^{-1} + \mathbf{I} \mathbf{B}_\mu^{-1})^{-1} \\ \mu_1 &= \Sigma_1 (\Sigma_0^{-1} \mu_0 + \mathbf{B}_\mu^{-1} \mathbf{S}_\mu). \end{aligned}$$

The structure of μ_1 and Σ_1 is logical. The precision of the full conditional is the sum of the precisions originating from the prior distribution and the likelihood function, while the mean of the full conditional is a weighted average of the prior mean and $\bar{\mu}$, the sample mean of the individual effects μ_i ($\mathbf{S}_\mu = I \times \bar{\mu}$). A parallel strategy is followed for obtaining full conditionals for the other hierarchical mean vectors α_ϕ , α_Γ and α_Δ . For each sample of α_ϕ , the stationarity condition $I(|\lambda(\mathbf{F}_i)| < 1)$ is checked (as explained in Eq. (12)). If the condition is not satisfied, we reject and resample.

A.3. Full conditionals for individual effect parameters

Finally, full conditional distributions are derived for the individual effects μ_i , $\Phi_{1,i}, \dots, \Phi_{L,i}$, Γ_i and Δ_i . Although this technique strongly resembles the one used for the hierarchical mean vectors as explained in the previous paragraphs, there are some complicating factors that need some extra calculations and transformations. Before we combine the likelihood and the prior distribution, we correct the data or states for non-relevant constants in the model and subsequently we define a vectorized model that takes into account all measurements at once. We will illustrate this approach for Γ_i , the matrix of regression effects for the observation covariates $X_{i,t}$.

As the focal variable is Γ_i , only the observation equation is relevant for the likelihood function and thus the transition equation can be ignored for now. To keep the notation simple, we group non-relevant constants (we also drop the conditioning notation at the left-hand side of the equation):

$$\begin{aligned} y_{t,i} &\sim N(\mu_i + \theta_{t,i} + \Gamma_i x_{t,i}, \Sigma_\epsilon) \\ y_{t,i} - \mu_i - \theta_{t,i} &\sim N(\Gamma_i x_{t,i}, \Sigma_\epsilon) \\ y_{t,i,\text{cor}} &\sim N(\Gamma_i x_{t,i}, \Sigma_\epsilon), \end{aligned} \quad (19)$$

where the subscript cor indicates that the observation vector $y_{t,i}$ was corrected for μ_i and $\theta_{t,i}$, which are both assumed to be known in this stage of the Gibbs sampler. The likelihood function is based on Eq. (19). However, continuing to consider Γ_i as a matrix will bring us into trouble later. Therefore, we have to reparameterize Eq. (19) in such a way that Γ_i is expressed as a vector. The solution lies in considering the $P \times T$ (corrected) observation matrix $Y_{i,\text{cor}}$ instead of the P -dimensional (corrected) observation vectors $y_{t,i,\text{cor}}$ for $t = 1, \dots, T$, which are in fact columns in $Y_{i,\text{cor}}$. By vectorizing this matrix with the $\text{vec}(\cdot)$ operator, we obtain the following model:

$$\text{vec}(Y_{i,\text{cor}}) \sim N(\text{vec}(X_i' \Gamma_i'), \Sigma_\epsilon \otimes I_T), \quad (20)$$

with X_i being the $J \times T$ matrix with the J -dimensional observation covariate vectors $x_{t,i}$ for $t = 1, \dots, T$ as its columns. A Kronecker product (\otimes) was used to expand the covariance matrix Σ_ϵ over all PT elements of $\text{vec}(Y_{i,\text{cor}})$. Using a basic property of vectorization and the Kronecker product, the vectorized mean term can be further decomposed as follows:

$$\text{vec}(Y_{i,\text{cor}}) \sim N((I_P \otimes X_i') \text{vec}(\Gamma_i'), \Sigma_\epsilon \otimes I_T). \quad (21)$$

For notational convenience, we use the $*$ -notation, indicating that we work in the vectorized model.

$$Y_{i,\text{cor}}^* \sim N(X_i^* \Gamma_i^*, \Sigma_\epsilon^*), \quad (22)$$

with $Y_{i,\text{cor}}^* = \text{vec}(Y_{i,\text{cor}})$, $X_i^* = (I_P \otimes X_i')$, $\Gamma_i^* = \text{vec}(\Gamma_i')$ and $\Sigma_\epsilon^* = \Sigma_\epsilon \otimes I_T$. The likelihood of Γ_i^* can now be expressed as a multivariate normal. Moreover, the hierarchical distribution defined in Eq. (10) serves as an informative multivariate normal prior distribution for $\text{vec}(\Gamma_i') = \Gamma_i^*$. They are combined into a multivariate normal full conditional distribution:

$$\begin{aligned} P(\Gamma_i^* | D, P) &\propto P(\Gamma_i^*) \ell(\Gamma_i^*) \\ &\propto N(\Gamma_i^* | \mu_0, \Sigma_0) \times N(Y_{i,\text{cor}}^* | X_i^* \Gamma_i^*, \Sigma_\epsilon^*) \\ &\dots \\ &\sim N(\mu_1, \Sigma_1) \end{aligned} \quad (23)$$

with

$$\begin{aligned} \Sigma_1 &= (\Sigma_0^{-1} + X_i^{*'} \Sigma_\epsilon^{*-1} X_i^*)^{-1} \\ \mu_1 &= \Sigma_1 (\Sigma_0^{-1} \mu_0 + X_i^{*'} \Sigma_\epsilon^{*-1} Y_{i,\text{cor}}^*). \end{aligned}$$

We do not report all steps in the derivation because the procedure is parallel with the technique discussed for the hierarchical mean vectors in Eq. (18). Returning from the $*$ -notation to the original notation, we use properties of the Kronecker product and vectorization:

$$\begin{aligned} \Sigma_1 &= (\Sigma_0^{-1} + \Sigma_\epsilon^{-1} \otimes (X_i X_i'))^{-1} \\ \mu_1 &= \Sigma_1 (\Sigma_0^{-1} \mu_0 + \text{vec}(X_i Y_{i,\text{cor}}' \Sigma_\epsilon^{-1})). \end{aligned}$$

To sample Γ_i , we sample Γ_i^* from its multivariate normal full conditional and transform it back to its matrix form Γ_i . The procedure is parallel for deriving full conditionals for μ_i , $\Phi_{l,i}$ (for $l = 1, \dots, L$) and Δ_i , with the only difference that μ_i is already a vector and the vectorization step can be skipped. Moreover, the stationarity condition is checked for the $\Phi_{l,i}$ (for $l = 1, \dots, L$), as explained in Eq. (12).

Appendix B. Parameter recovery

To evaluate the quality of the Gibbs sampler routine that we implemented in *R*, we report a small simulation study that shows the parameter recovery of one simulated set of observations. We explain the process of data simulation and parameter estimation and report our findings. It is not our goal to set up an intensive simulation study to investigate specific properties of the algorithm. We only want to illustrate that the algorithm obtains estimated values that are close to the population values.

B.1. Data simulation

The data simulation was based on the results of the hierarchical analysis for the depressed population of 67 individuals (see the application section). In a first step, we sampled the individual level parameters from hierarchical normal distributions. The estimated¹⁴ hierarchical mean vectors $\hat{\alpha}_\mu$, $\hat{\alpha}_\phi$ and $\hat{\alpha}_\Delta$ and hierarchical covariance matrices \hat{B}_μ , \hat{B}_ϕ and \hat{B}_Δ were used to define hierarchical Gaussian sampling distributions (see Table B.1).

After sampling from the hierarchical parameter distributions, a bivariate series y_i^{sim} of $T = 60$ observations was simulated for each of 67 individuals, using parameter sets $\{\tilde{\mu}_i, \tilde{\Phi}_{1,i}, \tilde{\Phi}_{2,i}, \tilde{\Phi}_{3,i}, \tilde{\Phi}_{4,i}, \tilde{\Phi}_{5,i}, \tilde{\Delta}_i, \Sigma_\epsilon, \Sigma_\eta\}$ and a covariate z_i^{sim} to account for the regression effect $\tilde{\Delta}_i$ in the transition equation. The tilde (\sim) indicates that parameters have been sampled from the hierarchical distributions. The values of Σ_ϵ and Σ_η were set to the estimated values in the application (see Table B.1). The values for covariate z_i^{sim} were simulated for each individual i from a standard normal distribution. The data, consisting of y_i^{sim} and z_i^{sim} for 67 individuals, are available as an *R* workspace at <http://sites.google.com/site/tomlodewyckx/downloads>.

B.2. Model estimation

With the simulated observations y_i^{sim} and covariate z_i^{sim} , we estimated the hierarchical state space model as described in the application in Eq. (14): a bivariate state-space model of order five, with a regression effect in the transition equation. We estimated three Markov chains of 5000 values, after removing a burn-in period of 1000 values.

B.3. Results of the simulated data

Convergence is controlled with \hat{R} and very similar values are found as for the analysis of the data from the depressed individuals. The estimated posterior quantiles for the parameters, using

¹⁴ The term “estimate” refers in this paragraph to the posterior median.

Table B.1

The left side of the table contains the estimated posterior quantiles $q_{0.025}$, $q_{0.500}$, $q_{0.975}$ (obtained with the Oregon data), which were used as population values for the data simulation. At the right side of the table, the estimated posterior quantiles are listed for the model identical model that was applied on the simulated data.

Description		Depressed			Simulated		
		$q_{0.025}$	$q_{0.500}$	$q_{0.975}$	$q_{0.025}$	$q_{0.500}$	$q_{0.975}$
<i>Hierarchical parameters μ</i>							
α_{μ_1}	HM for general mean HR	76.19	78.96	81.69	76.22	79.11	82.02
α_{μ_2}	HM for general mean BP	88.67	91.82	94.88	91.52	95.52	99.36
β_{μ_1}	HV for general mean HR	91.33	127.71	184.43	96.32	140.57	211.22
β_{μ_2}	HV for general mean BP	113.90	158.09	229.16	191.16	266.25	384.84
<i>Hierarchical parameters Φ_1</i>							
$\alpha_{\phi_{1,1 1}}$	HM for AR effect HR	0.85	0.93	1.00	0.59	0.66	0.73
$\alpha_{\phi_{1,2 1}}$	HM for CR effect BP → HR	−0.22	−0.17	−0.11	−0.17	−0.11	−0.04
$\alpha_{\phi_{2,1 1}}$	HM for CR effect HR → BP	0.11	0.17	0.23	0.05	0.12	0.19
$\alpha_{\phi_{2,2 1}}$	HM for AR effect BP	0.33	0.40	0.47	0.27	0.34	0.41
$\beta_{\phi_{1,1 1}}$	HV for AR effect HR	0.03	0.04	0.07	0.03	0.05	0.07
$\beta_{\phi_{1,2 1}}$	HV for CR effect BP → HR	0.02	0.04	0.05	0.03	0.05	0.08
$\beta_{\phi_{2,1 1}}$	HV for CR effect HR → BP	0.02	0.04	0.06	0.04	0.06	0.09
$\beta_{\phi_{2,2 1}}$	HV for AR effect BP	0.04	0.06	0.09	0.04	0.05	0.08
<i>Hierarchical parameters Δ</i>							
α_{δ_1}	HM for effect AP → HR	−0.88	−0.38	0.13	−0.46	−0.23	−0.01
α_{δ_2}	HM for effect AP → BP	−0.12	0.38	0.86	0.19	0.39	0.59
β_{δ_1}	HV for effect AP → HR	0.30	1.08	3.49	0.24	0.46	0.82
β_{δ_2}	HV for effect AP → BP	0.13	0.41	1.26	0.10	0.22	0.45
<i>Error variances</i>							
$\sigma_{\epsilon,1}^2$	Measurement error variance HR	0.49	0.82	1.30	2.26	3.19	4.36
$\sigma_{\epsilon,2}^2$	Measurement error variance BP	0.12	0.40	0.99	5.33	7.08	9.25
$\sigma_{\eta,1}^2$	Innovation variance HR	14.63	15.47	16.39	18.57	19.91	21.35
$\sigma_{\eta,2}^2$	Innovation variance BP	21.57	22.65	23.78	21.19	22.70	24.27

the simulated data, are reported in Table B.1. In general, we see that the parameter values approximate the population values. There are some remarkable differences between the population values and the ones based on the simulation. For instance, the dynamic parameters in α_{ϕ_1} and the regression effects in α_{Δ} are lower than in the population, which is compensated for with an overestimation of the error variances in Σ_{ϵ} and Σ_{η} .

Despite these differences, the general trend is more or less the same. Increasing the amount of observation moments T and individuals N should bring the estimates closer to the population values. An extensive simulation study, varying the population values, should reveal more information on how the algorithm performs.

References

Allen, N. B., & Badcock, P. B. T. (2003). The social risk hypothesis of depressed mood: evolutionary, psychosocial, and neurobiological perspectives. *Psychological Bulletin*, 129, 887–913.

Allen, N. B., & Sheeber, L. (2008). *Adolescent emotional development and the emergence of depressive disorders*. Cambridge, UK: Cambridge University Press.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, 58, 373–403.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Borrelli, R., & Coleman, C. (1998). *Differential equations: a modeling perspective*. New York: Wiley.

Bosworth, H. B., Bartash, R. M., Olsen, M. K., & Steffens, D. C. (2003). The association of psychosocial factors and depression with hypertension among older adults. *International Journal of Geriatric Psychiatry*, 18, 1142–1148.

Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 581–607). New York: Cambridge University Press.

Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review*, 28, 676–691.

Byrne, M. L., Sheeber, L., Simmons, J. G., Davis, B., Shortt, J. W., & Katz, L. F. et al. Autonomic cardiac control in depressed adolescents. *Depression & Anxiety* (in press).

Carney, R. M., Freedland, K. E., & Veith, R. C. (2005). Depression, the autonomic nervous system, and coronary heart disease. *Psychosomatic Medicine*, 67, S29–S33.

Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81, 541–553.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46, 167–174.

Cook, E., & Turpin, G. (1997). Differentiating orienting, startle, and defense responses: the role of affect and its implications for psychopathology. In *Attention and orienting: sensory and motivational processes*.

Darwin, C., Ekman, P., & Prodger, P. (2002). *The expression of the emotions in man and animals*. Oxford: Oxford University Press.

Davidson, R. J. (1998). Affective style and affective disorders: perspectives from affective neuroscience. *Cognition & Emotion*, 12, 307–330.

Davidson, K., Jonas, B. S., Dixon, K. E., & Markovitz, J. H. (2000). Do depression symptoms predict early hypertension incidence in young adults in the CARDIA study? *Archives of Internal Medicine*, 160, 1495–1500.

Dawson, M. E., Schell, A. M., & Catania, J. J. (1977). Autonomic correlates of depression and clinical improvement following electroconvulsive shock therapy. *Psychophysiology*, 14, 569–578.

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford University Press.

Frasure-Smith, N., Lesperance, F., Irwin, M. R., Talajic, M., & Pollock, B. G. (2009). The relationships among heart rate variability, inflammatory markers and depression in coronary heart disease patients. *Brain, Behavior, and Immunity*, 23, 1140–1147.

Frijda, N. H., Mesquita, B., Sonnemans, J., & Van Goozen, S. (1991). The duration of affective phenomena or emotions, sentiments and passions. In K. Strongman (Ed.), *International review of studies on emotion: Vol. 1* (pp. 187–225). UK: Wiley.

Gangwisch, J. E., Malaspina, D., Posner, K., Babiss, L. A., Heymsfield, S. B., Turner, J. B., et al. (2010). Insomnia and sleep duration as mediators of the relationship between depression and hypertension incidence. *American Journal of Hypertension*, 23, 62–69.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (second ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markovchain Monte Carlo in practice*. Chapman & Hall/CRC.

Grewal, M. S., & Andrews, A. P. (2001). *Kalman filtering: theory and practice using MATLAB* (second ed.). New York: Wiley.

Gross, J. J. (2002). Emotion regulation: affective, cognitive and social consequences. *Psychophysiology*, 39, 281–291.

Gupta, R. K. (2009). Major depression: an illness with objective physical signs. *World Journal of Biological Psychiatry*, 10, 196–201.

Hamilton, J. D. (1994). *Time series analysis*. New Jersey: Princeton University Press.

Hojtink, H., Klugkist, I., & Boelen, P. (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.

- Hops, H., Davis, B., & Longoria, N. (1995). Methodological issues in direct observation—illustrations with the living in familial environments (LIFE) coding system. *Journal of Clinical Child Psychology*, 2, 193–203.
- Hughes, J. W., Casey, E., Luster, F., Doe, V. H., Waechter, D., Rosneck, R. N., et al. (2006). Depression symptoms predict heart rate recovery after treadmill stress testing. *American Heart Journal*, 151, 1122.e1–1122.e6.
- Jonas, B. S., Franks, P., & Ingram, D. D. (1997). Are symptoms of anxiety and depression risk factors for hypertension? Longitudinal evidence from the national health and nutrition examination survey I epidemiologic follow-up study. *Archives of Family Medicine*, 6, 43–49.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 35–45.
- Kim, C. J., & Nelson, C. R. (1999). *State-space models with regime switching*. Cambridge, MA: MIT Press.
- Klugkist, I., & Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models. *Computational Statistics and Data Analysis*, 51, 6367–6379.
- Kuppens, P., Allen, N. B., & Sheeber, L. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, 21, 984–991.
- Kuppens, P., Stouten, J., & Mesquita, B. (2009). Individual differences in emotion components and dynamics. *Cognition & Emotion*, 23, 373–403. (special issue).
- Lahmeyer, H. W., & Bellur, S. N. (1987). Cardiac regulation and depression. *Journal of Psychiatric Research*, 21, 1–6.
- Lake, C. R., Pickar, D., Ziegler, M. G., Lipper, S., Slater, S., & Murphy, D. L. (1982). High plasma norepinephrine levels in patients with major affective disorder. *American Journal of Psychiatry*, 139, 1315–1318.
- Lauritzen, S. L. (1981). Time series analysis in 1880: a discussion of contributions made by T.N. Thiele. *International Statistical Review*, 49, 319–331.
- Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin and Review*, 12, 605–621.
- Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science*, 3, 23–27.
- Licht, C. M. M., De Geus, E. J. C., Seldenrijk, A., van Hout, H. P. J., Zitman, F. G., van Dyck, R., et al. (2009). Depression is associated with decreased blood pressure, but antidepressant use increases the risk for hypertension. *Hypertension*, 53, 631–638.
- Licht, C. M. M., De Geus, E. J. C., Zitman, F. G., Hoogendijk, W. J. G., van Dyck, R., & Penninx, B. W. J. H. (2008). Association between major depressive disorder and heart rate variability in the Netherlands study of depression and anxiety (NESDA). *Archives of General Psychiatry*, 65, 1358–1367.
- Liu, D., Niu, X. F., & Wu, H. (2004). *Mixed-effects state space models for analysis of longitudinal data*.
- Ljung, L. (1999). *System identification-theory for the user* (second ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). Winbugs—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325–337.
- Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics* (second ed.). New York: Wiley.
- Matsukawa, S., & Wada, T. (1997). Vector autoregressive modeling for analyzing feedback regulation between heart rate and blood pressure. *American Journal of Physiology—Heart and Circulatory Physiology*, 273, 478–486.
- Meinhold, R. J., & Singpurwalla, N. D. (1983). Understanding the Kalman filter. *The American Statistician*, 37, 123–127.
- Merkle, E., Smithson, M., & Verkuilen, J. (2011). Hierarchical models of simple mechanisms underlying confidence in decision making. *Journal of Mathematical Psychology*, 55(1), 57–67.
- Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55(1), 84–93.
- Nishiyama, S., Yana, K., Mizuta, H., & Ono, T. (2007). Effect of the presence of long term correlated common source fluctuation to heart rate feedback analysis. In *Noise and Fluctuations—AIP Conference Proceedings*. Vol. 922 (pp. 695–698).
- Oatley, K., & Jenkins, J. M. (1992). Human emotion—function and dysfunction. *Annual Review of Psychology*, 43, 55–85.
- Orvaschel, H., & Puig-Antich, J. (1994). Schedule of affective disorders and schizophrenia for school-age children: epidemiologic version. *Unpublished manual*.
- Panerai, R. B., James, M. A., & Potter, J. F. (1997). Impulse response analysis of baroreceptor sensitivity. *American Journal of Physiology—Heart and Circulatory Physiology*, 272, 1866–1875.
- Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.
- Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic linear models with R*. New York: Springer.
- Pizzi, C., Manzoli, L., Mancini, S., & Costa, G. M. (2008). Analysis of potential predictors of depression among coronary heart disease risk factors including heart rate variability, markers of inflammation, and endothelial function. *European Heart Journal*, 29, 1110–1117.
- Pourahmadi, M., & Daniels, M. J. (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics*, 58, 225–231.
- Pratte, M., & Rouder, J. (2011). Hierarchical single- and dual-process models of recognition memory. *Journal of Mathematical Psychology*, 55(1), 36–46.
- R Development Core Team (2009). R: a language and environment for statistical computing. Vienna, Austria. ISBN: 3-900051-07-0.
- Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385–401.
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge: Blackwells.
- Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods* (second ed.). New York: Springer-Verlag.
- Rottenberg, J. (2005). Mood and emotion in major depression. *Current Directions in Psychological Science*, 14, 167–170.
- Rutledge, T., & Hogan, B. E. (2002). A quantitative review of prospective evidence linking psychological factors with hypertension development. *Psychosomatic Medicine*, 64, 758–766.
- Scherer, K. R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In M. Lewis, & I. Granic (Eds.), *Emotion, development, and selforganization: dynamic systems approaches to emotional development* (pp. 70–99). New York: Cambridge University Press.
- Scherer, K. R. (2009). The dynamic architecture of emotion: evidence for the component process model. *Cognition & Emotion*, 23, 1307–1351.
- Scherrer, J. F., Xian, H., Bucholz, K. K., Eisen, S. A., Lyons, M. J., Goldberg, J., et al. (2003). A twin study of depression symptoms, hypertension, and heart disease in middle-aged men. *Psychosomatic Medicine*, 65, 548–557.
- Sheeber, L., Allen, N. B., Leve, C., Davis, B., Shortt, J. W., & Katz, L. F. (2009). Dynamics of affective experience and behavior in depressed adolescents. *Journal of Child Psychology and Psychiatry*, 50, 1419–1427.
- Sheeber, L. B., Davis, B., Leve, C., Hops, H., & Tildesley, E. (2007). Adolescents' relationships with their mothers and fathers: associations with depressive disorder and subdiagnostic symptomatology. *Journal of Abnormal Psychology*, 116, 144–154.
- Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32, 1248–1284.
- Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications: with R examples* (second ed.). New York: Springer.
- Simon, D. (2006). *Optimal state estimation: Kalman, H_∞ and nonlinear approaches*. Hoboken, NJ: Wiley-Interscience.
- Sparrenberger, F., Cichelerio, F. T., Ascoli, A. M., Fonseca, F. P., Weiss, G., Berwanger, O., et al. (2006). Does psychosocial stress cause hypertension? A systematic review of observational studies. *Journal of Human Hypertension*, 20, 434–439.
- Triedman, J. K., Perrott, M. H., Cohen, R. J., & Saul, J. P. (1995). Respiratory sinus arrhythmia—time-domain characterization using autoregressive moving average analysis. *American Journal of Physiology—Heart and Circulatory Physiology*, 268, 2232–2238.
- Van Overschee, P., & De Moor, B. (1996). *Subspace identification for linear systems: theory, implementations, applications*. Kluwer Academic Publishers.
- Wiehe, M., Fuchs, S. C., Moreira, L. B., Moraes, R. S., Pereira, G. M., Gus, M., et al. (2006). Absence of association between depression and hypertension: results of a prospectively designed population-based study. *Journal of Human Hypertension*, 20, 434–439.